GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:40:43 ; Search time 69.875 Seconds
(without alignments)
3923.374 Million cell updates/sec

Title: US-09-000-062-7
Perfect score: 494
Sequence: 1 ctcaggcgaagaacaggtat.....cgcagatcccgggatctgcg 494

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	494	100.0	494	4	US-09-000-062-7	Sequence 7, Appli
	2	49.6	10.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	3	38.2	7.7	6669	4	US-10-204-708-5	Sequence 5, Appli
	4	37.2	7.5	832	4	US-09-621-976-2813	Sequence 2813, Ap
	5	37.2	7.5	1230025	4	US-09-198-452A-1	Sequence 1, Appli
c	6	37	7.5	1193	4	US-09-372-422A-23	Sequence 23, Appl
	7	36.2	7.3	5562	4	US-10-204-708-63	Sequence 63, Appl
	8	36.2	7.3	6866	4	US-10-204-708-20	Sequence 20, Appl
c	9	35.8	7.2	1493	1	US-08-340-820-24	Sequence 24, Appl
c	10	35.8	7.2	1493	1	US-08-593-535-24	Sequence 24, Appl
	11	35	7.1	10467	4	US-10-204-708-1	Sequence 1, Appli
c	12	34.8	7.0	289	3	US-09-007-005-17	Sequence 17, Appl
c	13	34.8	7.0	289	3	US-09-244-796-17	Sequence 17, Appl
	14	34.4	7.0	10467	4	US-10-204-708-2	Sequence 2, Appli
c	15	34.2	6.9	1406	3	US-08-913-842-6	Sequence 6, Appli
	16	34	6.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
c	17	33.8	6.8	1051	4	US-09-396-149-15	Sequence 15, Appl
c	18	33.6	6.8	531	4	US-09-328-352-2881	Sequence 2881, Ap
	19	33.6	6.8	7304	4	US-10-204-708-43	Sequence 43, Appl
	20	33.6	6.8	11015	4	US-10-204-708-55	Sequence 55, Appl
c	21	33.2	6.7	1896	4	US-09-107-532A-894	Sequence 894, App
	22	32.8	6.6	5666	4	US-10-204-708-29	Sequence 29, Appl
	23	32.8	6.6	8961	4	US-10-204-708-80	Sequence 80, Appl
c	24	32.8	6.6	15071	4	US-09-358-082A-29	Sequence 29, Appl
	25	32.6	6.6	11014	4	US-08-956-171E-91	Sequence 91, Appl
c	26	32.4	6.6	1545	4	US-09-107-532A-2628	Sequence 2628, Ap
	27	32.4	6.6	5360	4	US-10-204-708-65	Sequence 65, Appl
	28	32.2	6.5	8537	4	US-10-204-708-41	Sequence 41, Appl
	29	32.2	6.5	8607	4	US-10-204-708-71	Sequence 71, Appl
c	30	32	6.5	2022	4	US-08-956-171E-244	Sequence 244, App
	31	32	6.5	46718	4	US-09-816-093-3	Sequence 3, Appli
	32	31.8	6.4	476	4	US-09-621-976-10903	Sequence 10903, A
c	33	31.8	6.4	4998	4	US-09-453-702B-21	Sequence 21, Appl
	34	31.8	6.4	6070	4	US-10-204-708-10	Sequence 10, Appl
	35	31.8	6.4	6113	4	US-10-204-708-14	Sequence 14, Appl
c	36	31.6	6.4	789	4	US-09-134-001C-400	Sequence 400, App
c	37	31.6	6.4	1453	4	US-09-976-594-760	Sequence 760, App
	38	31.6	6.4	2301	4	US-09-134-001C-852	Sequence 852, App
	39	31.6	6.4	4072	3	US-09-272-496-7	Sequence 7, Appli
	40	31.6	6.4	11126	4	US-08-956-171E-171	Sequence 171, App
	41	31.6	6.4	11131	4	US-10-204-708-27	Sequence 27, Appl
c	42	31.6	6.4	41100	4	US-09-755-665-46	Sequence 46, Appl
	43	31.4	6.4	1884	1	US-08-307-382-3	Sequence 3, Appli
	44	31.4	6.4	1884	1	US-08-366-779-3	Sequence 3, Appli
	45	31.4	6.4	1884	1	US-08-478-727-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-000-062-7
; Sequence 7, Application US/09000062
; Patent No. 6338961
; GENERAL INFORMATION:
; APPLICANT: DEROSE, Richard
; APPLICANT: CHAUBET, Nicole
; APPLICANT: GIGOT, Claude (deceased)
; TITLE OF INVENTION: ISOLATED DNA SEQUENCE CAPABLE OF SERVING AS REGULATORY
; TITLE OF INVENTION: ELEMENT IN A CHIMERIC GENE WHICH CAN BE USED FOR THE
; TITLE OF INVENTION: TRANSFORMATION OF PLANTS
; FILE REFERENCE: 022650-453
; CURRENT APPLICATION NUMBER: US/09/000,062
; CURRENT FILING DATE: 1998-05-29
; EARLIER APPLICATION NUMBER: PCT/FR96/01109
; EARLIER FILING DATE: 1996-07-17
; EARLIER APPLICATION NUMBER: FR 95/08980
; EARLIER FILING DATE: 1995-07-19
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 494
; TYPE: DNA
; ORGANISM: Zea mays
US-09-000-062-7

Query Match 100.0%; Score 494; DB 4; Length 494;
Best Local Similarity 100.0%; Pred. No. 2e-128;
Matches 494; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
```

```
Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480

Qy        481 TCCCGGGATCTGCG 494
              ||||||||||||||
Db        481 TCCCGGGATCTGCG 494
```

RESULT 2
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 10.0%; Score 49.6; DB 1; Length 7218;
Best Local Similarity 6.1%; Pred. No. 0.00022;
Matches 25; Conservative 213; Mismatches 172; Indels 0; Gaps 0;

```
Qy         14 CAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGT 73
              |||||     | |||  ||:   ::    :::   ::::::: :: ::::::  : :
Db       1046 CAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1105
```

```
Qy         74 TTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCC 133
              :::::: :: :: ::::: :  :::  :::: :  :  :    ::::   :::: ::::
Db       1106 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1165

Qy        134 ATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAG 193
              ::::  :::     :  : ::: ::::::  ::: ::  :  :   :  ::: :
Db       1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225

Qy        194 ATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAG 253
              ::   :::  ::  ::: ::   :  ::: :::::: :: :::::::::: ::::
Db       1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285

Qy        254 TTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACT 313
              :::: : :  ::   ::: ::: : : :  :: :: :: :::  :: : : ::     ::
Db       1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345

Qy        314 AGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTG 373
                  :::::   ::  :::: :::::::  :: : :::  :::  :  :  ::: :::
Db       1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405

Qy        374 GATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTT 423
              ::    ::::   :  :  : :::  : | |||| |||||| |  | |
Db       1406 YYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCT 1455
```

RESULT 3
US-10-204-708-5
; Sequence 5, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 5
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-5

Query Match 7.7%; Score 38.2; DB 4; Length 6669;
Best Local Similarity 48.0%; Pred. No. 0.32;
Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
Qy         50 GGTCTTTCCATTACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGA 109
              ||| |||   ||  |||||  | | |||| | || |  |  |  | | | ||  |
Db       5179 GGTTTTTGATTTTTTTTTTTTTTTATTTTATTTTTTTTGGTTTTTTTTCGGGTGTTTAGC 5238

Qy        110 CAATAGAGTTTCGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTT 169
              ||   |||      |||| |    ||  |  ||   |||  |  |  || ||||  ||
Db       5239 GGATTTGGTTAATAATTTTTTTTTTTTAAACGTTTTGAATTGAATTTTGTTTTTTGCGTA 5298

Qy        170 TGCTGGATCGATAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTT 229
              | |   |  |    | |  | ||  | || |      ||| ||| |  ||| ||
Db       5299 TTTTTTTTTTGTGTTATTTTAGGGTTTAGATTTAATTACGCGATTTGATTAGTTATTTTT 5358
```

```
Qy        230 CATCGTTTTTTTCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTG 276
              ||||||||  |  || | | || |   || | ||  ||   ||
Db       5359 TGATTTTTTTTTCGTATTTAGTATTTTTAGTTAGCGAGTATTTATTG 5405


RESULT 4
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             7.5%;  Score 37.2;  DB 4;  Length 832;
  Best Local Similarity   9.6%;  Pred. No. 0.31;
  Matches   35;  Conservative  176;  Mismatches  154;  Indels    1;  Gaps    1;

Qy         71 TGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGT 130
              :::::::: | :  |: : : : : : :::::: : : : ::: :  :::: : :|:: :
Db          1 YRWYWWKYTTWYAKCWTKWKWSWSYWMYWKWYYMKTYWRWRRKKKKAWWKYWKTWTWWYW 60

Qy        131 CCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTG 190
              : ::  : ::  :   :    : ::   :: :::::    |: :|  |: : ::: ::
Db         61 RYAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYMAMWTRTWTGYAYYRSMMYWWR 120

Qy        191 AAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCG 250
              : ::::| : :   :: : : :::::::|:  ::: :|: :  | :::  |:|: :  :
Db        121 YRCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKTYYWAATRYWWMMCWTKRWRAS 180

Qy        251 AAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGA 310
              ::  : :|:|  :: ::      : :  ::|    :: |  : : ::: ::: :  |||:
Db        181 WWYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGAMSWKYMWRMWRWRGWATGAGM 240

Qy        311 ACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCAT 370
              :: :       ||::   ::   :| |  ::  : :|  || : ::   : |    ::
Db        241 KAWRASCMMRRKYAGKSKTSYKSMWMCWTRSWKYCYTKARWTGYYCYRKGGMWGKRGRWY 300

Qy        371 -CTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTAT 429
               ::: ::::    :  : :::   : : ::::: :|::: :::: : :||: ::| ::|:
Db        301 ASKKYMWKRWWWCWARMYRYSTGTRASMWWRRWYYTMMMKWWKYAWARAAWRWWAMWWAW 360

Qy        430 GGATAA 435
              ::| ||
Db        361 RRACAA 366


RESULT 5
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
;   LENGTH: 1230025
;   TYPE: DNA
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(15000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (15001)..(30000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (30001)..(45000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (45001)..(60000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (60001)..(75000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (75001)..(90000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (90001)..(105000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (105001)..(120000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (120001)..(135000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (135001)..(150000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (150001)..(165000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (165001)..(180000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (180001)..(195000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (195001)..(210000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (210001)..(225000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (225001)..(240000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (240001)..(255000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (255001)..(270000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (270001)..(285000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (285001)..(300000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (300001)..(315000)
;   OTHER INFORMATION: n=a or c or g or t
;   NAME/KEY: misc_feature
;   LOCATION: (315001)..(330000)
;   OTHER INFORMATION: n=a or c or g or t
```

```
;    NAME/KEY: misc_feature
;    LOCATION: (330001)..(345000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (345001)..(360000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (360001)..(375000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (375001)..(390000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (390001)..(405000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (405001)..(420000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (420001)..(435000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (435001)..(450000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (450001)..(465000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (465001)..(480000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (480001)..(495000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (495001)..(510000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (510001)..(525000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (525001)..(540000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (540001)..(555000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (555001)..(570000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (570001)..(585000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (585001)..(600000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (600001)..(615000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (615001)..(630000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (630001)..(645000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (645001)..(660000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (660001)..(675000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (675001)..(690000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (690001)..(705000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (705001)..(720000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (720001)..(735000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (735001)..(750000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (750001)..(765000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (765001)..(780000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (780001)..(795000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (795001)..(810000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (810001)..(825000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (825001)..(840000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (840001)..(855000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (855001)..(870000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (870001)..(885000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (885001)..(900000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature
;    LOCATION: (900001)..(915000)
;    OTHER INFORMATION: n=a or c or g or t
;    NAME/KEY: misc_feature

  Query Match             7.5%;  Score 37.2;  DB 4;  Length 1230025;
  Best Local Similarity   44.3%;  Pred. No. 3.3;
  Matches  141;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy         34 TTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGTTACTGTCTCCG 93
              | ||   || |    | | |  ||   || |||   ||    |||     |  |||||
Db    1129940 TGAGTAGAGAGAAATTGTTNATTTTGTTTCTTNNCAAANATTTTATCTTACCAGTCTCTT 1129999

Qy         94 CGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTTTGAAAATAAAC 153
              | || ||  |         ||    || |  | |  || |  |||    | |||    |
Db    1130000 CTATTTGTCTGAGTTACTTGAGTGATGCGTGAATTTTTACATGTGACTTTTGGATCATGG 1130059

Qy        154 GTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCTAGTCGATTTAT 213
               |          | |    |||     |    | |  ||||       |       |||||
Db    1130060 ATAACTGGAACTTTTATAAAAGATATCGAGTTCAGAAAAGAAGAAAAAAACCTCTTTTAT 1130119

Qy        214 TGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATAACCAGATTTGT 273
               |  |||||  ||| || |   || | |    |      ||  || |       |||   |
Db    1130120 TTAATGATTTATTTTTTAAAGTATATAAATAATTTATAATTATAAAATAATTTACTTTTT 1130179

Qy        274 CTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTCGAGTCAATTTT 333
                          |   | |||  |       | ||    | |    ||  |  | |   |
Db    1130180 AATAATTAAAATGTAAACTGTTTGCACTTTAATAATAATTATAATTATAGATTTTGATGC 1130239

Qy        334 GCCCCTTTTGGTTATATC 351
                  || ||| ||  |||
Db    1130240 TTGTCTATTGTTTTGATC 1130257
```

```
RESULT 6
US-09-372-422A-23/c
; Sequence 23, Application US/09372422A
; Patent No. 6313375
; GENERAL INFORMATION:
;  APPLICANT: Rudolf Jung
;  APPLICANT:  Francois Barrieu
;  TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
;  FILE REFERENCE: 0919
;  CURRENT APPLICATION NUMBER: US/09/372,422A
;  CURRENT FILING DATE:  1999-08-11
;  PRIOR APPLICATION NUMBER: US 60/098,692
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 23
;   LENGTH: 1193
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (88)...(838)
US-09-372-422A-23

  Query Match             7.5%;  Score 37;  DB 4;  Length 1193;
  Best Local Similarity   53.9%;  Pred. No. 0.4;
  Matches   76;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db       1182 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1123

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db       1122 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTCGAAACGAAAATCACCTGTTTTTTCCCACTC 1063

Qy        138 CAGTTTGAAAATAAACGTCCG 158
Db       1062 CAAGGCGATCGAAAACGGCAG 1042


RESULT 7
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
;   LENGTH: 5562
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

  Query Match             7.3%;  Score 36.2;  DB 4;  Length 5562;
  Best Local Similarity   43.4%;  Pred. No. 1.1;
  Matches  167;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy         46 TTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTT 105
Db       1380 TTTTATTATTTTAATATTTTATTATTATTTTTTTTTTTTCGGTTAGGAATGATTTTTTT 1439

Qy        106 ACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTA 165
Db       1440 TTGATTATATTTATATTTTTTTATTTTAGTTTAGTTTTTTTTTTTTTTTATTTATTTTTTA 1499

Qy        166 AGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCCAT 225
Db       1500 TTATTTTTATTTTTTGTTAGTTAGATTTTTTTTTTTTTTATTATTTTTTTGTTTTTTATAG 1559

Qy        226 TCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATT 285
Db       1560 TGTTTTTTTTTATTTTTTTTTTGTTTAAGAAAGTTTATTAATTTTTATTTTTTTTATATT 1619

Qy        286 GTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGT 345
Db       1620 TTTTTTTTTTTTAGGATAGTGTTATTTTTAGAATTTTGTGATAATAATGATTTTTATTAT 1679

Qy        346 TATATCTGGTTCGATAACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATT 405
Db       1680 ATTATTTATTTTTGTTTTTTTTATTTTATGATGGTTTTTTTTTTTTTTTTGTTTTTTATT 1739

Qy        406 CCAATTTCTTCAAAATTTAGTTATG 430
Db       1740 GATTTTTTTTTACGTTTTCGTTTTG 1764


RESULT 8
US-10-204-708-20
; Sequence 20, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 20
;   LENGTH: 6866
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-20

  Query Match             7.3%;  Score 36.2;  DB 4;  Length 6866;
  Best Local Similarity   45.3%;  Pred. No. 1.2;
  Matches  208;  Conservative    0;  Mismatches  248;  Indels    3;  Gaps    2;

Qy         10 AGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTA 69
```

```
Db       5350 AGTAGAGTTGGGGTTTCGTTATGTTGGTTAGGTTGGTTTTGAATTTTTGATTTTAAGTAA 5409

Qy         70 ATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTG 129

Db       5410 TTTGTTTGTTTTAGTTTTTTAAAGTGTTGAGATTATAGGTATGAGTTATTATTTTTAGTT 5469

Qy        130 TCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGT 189

Db       5470 TAAAGTTTGTTTTTTATATTTTTTATGGTTTTTGATGTTTTTTTTGTTTTAAATTATTTT 5529

Qy        190 GAAGATTGAGTCTAGTCGATTTATTGGATGATCCA-TTCTTCATCGTTTTTTTCTTGCTT 248

Db       5530 TAATGAGGATTTTAATTATTTTATTAGAATTTAAATTTTTTTATTATTTTTTTTTTAATT 5589

Qy        249 CGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGA 308

Db       5590 TGGTTTTTTGTTTTGTATTTTTGTTTTTATTTTTATAGTTTTTTATTTTTGTTTTATAGA 5649

Qy        309 GAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTC 368

Db       5650 AGTTTTTTGATTTTATGTTTATTTTTTTATTTATTTTTATATAGATTTAGATATATGTAG 5709

Qy        369 ATCTGGATTAGGGT--TTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGT 426

Db       5710 ATTTTTTATATATTTATTTAGATATATTTAGATATAAATATATATTATTTTTATTTAAAT 5769

Qy        427 TATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCT 465

Db       5770 TATAGATATAGATAGATAGAAATTTATAGTTGGTTTTTT 5808


RESULT 9
US-08-340-820-24/c
; Sequence 24, Application US/08340820
; Patent No. 5512460
;  GENERAL INFORMATION:
;    APPLICANT:  NARUO, Ken-ichi
;    APPLICANT:  SEKO, Chisako
;    APPLICANT:  KUROKAWA, Tsutomu
;    APPLICANT:  KONDO, Tatsuya
;    TITLE OF INVENTION:  GLIA ACTIVATING FACTOR AND ITS
;    TITLE OF INVENTION:  PRODUCTION
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/340,820
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/835,713
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CONLIN, David G.
;      REGISTRATION NUMBER:  27026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      HAPLOTYPE:  2n
;      TISSUE TYPE:  skin
;      CELL TYPE:  fibroblast
;    IMMEDIATE SOURCE:
;      LIBRARY:  Human foreskin cDNA library
;      CLONE:  pGAF1
US-08-340-820-24

  Query Match             7.2%;  Score 35.8;  DB 1;  Length 1493;
  Best Local Similarity   50.3%;  Pred. No. 0.92;
  Matches   88;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77

Db       1482 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1423

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137

Db       1422 TTTTTTTGGAATTTCTATAAATAAATTTAACTTTTTTTATTTTTATTTTTATTTTTTTT 1363

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAA 192

Db       1362 TATTTTTTACTAAGAGGTCTTTGCTTTAAGTTCACTGCGATGATACCAGTGAGAA 1308


RESULT 10
US-08-593-535-24/c
; Sequence 24, Application US/08593535
; Patent No. 5622928
;  GENERAL INFORMATION:
;    APPLICANT:  NARUO, Ken-ichi
;    APPLICANT:  SEKO, Chisako
;    APPLICANT:  KUROKAWA, Tsutomu
;    APPLICANT:  KONDO, Tatsuya
;    TITLE OF INVENTION:  GLIA ACTIVATING FACTOR AND ITS
;    TITLE OF INVENTION:  PRODUCTION
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/593,535
;      FILING DATE:  24-JAN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/835,713
;      FILING DATE:  12-FEB-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CONLIN, David G.
;      REGISTRATION NUMBER:  27026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  24:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1493 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;      HAPLOTYPE:  2n
;      TISSUE TYPE:  skin
;      CELL TYPE:  fibroblast
;    IMMEDIATE SOURCE:
;      LIBRARY:  Human foreskin cDNA library
;      CLONE:  pGAF1
US-08-593-535-24

  Query Match             7.2%;  Score 35.8;  DB 1;  Length 1493;
  Best Local Similarity   50.3%;  Pred. No. 0.92;
  Matches   88;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db       1482 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1423

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db       1422 TTTTTTTGGAATTTCTATAAATAAATTTAACTTTTTTATTTTTATTTTTATTTTTTTT 1363

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAA 192
Db       1362 TATTTTTTACTAAGAGGTCTTTGCTTTAAGTTCACTGCGATGATACCAGTGAGAA 1308


RESULT 11
US-10-204-708-1
; Sequence 1, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 1
;   LENGTH: 10467
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-1

  Query Match             7.1%;  Score 35;  DB 4;  Length 10467;
  Best Local Similarity   44.4%;  Pred. No. 2.9;
  Matches  185;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    1;

Qy         20 TGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTC 79
Db      10045 TGTTTTGGTTATTTTAGGTTATTTTGTTTTTGTTTTTAAAATAATAATATATATTTGTTG 10104

Qy         80 TGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCA 139
Db      10105 TTTTGATTTTAAAGGGGTAATTGTTTTGGTAATTAGAAAATTATTTTTTTAATTTTTTTA 10164

Qy        140 GTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAG 199
Db      10165 AATTTTGTTAAAGGGAAAAGTAAGTTAGGTTGTTGGAGAGGTTA--TGTTGAGGTTTTAG 10222

Qy        200 TCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGT 259
Db      10223 AATTTTTGTATTTTTGGAAGTTTTGCGTGTTTTGTTTTTTGTTTTTTTTTTGTGTTTTT 10282

Qy        260 ATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTT 319
Db      10283 GTTGGTAGGTTTTTAGGTAGGATTTAGGAGGTAGGTAAGTTATTTTAGTTAAGTTATAAG 10342

Qy        320 TTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTAG 379
Db      10343 TTTATGGTTTAATTGTTTTTTTAGTTTTTTAAGGGTTGTTTTATAGGTAGTAAAGGGAGG 10402

Qy        380 GGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAAT 436
Db      10403 GGTTTTTATAGGTTTATTATTGTAGTTTTAATTTATTTTTTTTTTTTATTGTTTTAT 10459


RESULT 12
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

  Query Match             7.0%;  Score 34.8;  DB 3;  Length 289;
  Best Local Similarity   11.2%;  Pred. No. 1;
  Matches   27;  Conservative   86;  Mismatches  127;  Indels    0;  Gaps    0;

Qy         64 TTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGG 123
Db        286 TTTTTTTTTTTTTTTTTTTTTTTTTTTAYGYCYGYCYAYAYGYAYGYTYTYAYCYGYCYAY 227

Qy        124 GTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAA 183
Db        226 GYCYTYGYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNY 167

Qy        184 ACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCT 243
Db        166 NYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNY 107
```

```
Qy        244 TGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGT 303
                 :: :      : ::: :  ::    : : : ::: : ::  : ::  : :|:|  : :|:
Db        106 NYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYCYAYTYTYGY 47

RESULT 13
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

  Query Match              7.0%;  Score 34.8;  DB 3;  Length 289;
  Best Local Similarity   11.2%;  Pred. No. 1;
  Matches   27;  Conservative   86;  Mismatches  127;  Indels    0;  Gaps    0;

Qy         64 TTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGG 123
              |||||  | ||||||  | ||   | |   :   |  :|:    : :|:|     :|     : :
Db        286 TTTTTTTTTTTTTTTTTTTTTTTTTTAYGYCYGYCYAYAYGYAYGYTYTYAYCYGYCYAY 227

Qy        124 GTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAA 183
              |: :|   ::: : ::   : ::      :     : ::: : :        :     ::   :    :
Db        226 GYCYTYGYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNY 167

Qy        184 ACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCT 243
                :: :        :     : ::     : : :    :       ::: : : :    ::: : : :
Db        166 NYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNY 107

Qy        244 TGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGT 303
                 :: :      : ::: :  ::    : : : ::: : ::  : ::  : :|:|  : :|:
Db        106 NYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYCYAYTYTYGY 47

RESULT 14
US-10-204-708-2
; Sequence 2, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 2
;   LENGTH: 10467
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-2

  Query Match              7.0%;  Score 34.4;  DB 4;  Length 10467;
  Best Local Similarity   45.1%;  Pred. No. 4.2;
  Matches  128;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy          8 GAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTT 67
              |||  | |     ||   |||  |||   | ||||  |         |||      |||   |     |
Db       2334 GAATTAAATTAATAGTTTTTTAAATTTAGTTTTTTTTTTGTAGTTTTTATTGTTAATG 2393

Qy         68 TAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTT 127
                ||  |||||  | |||  |  |  |        | |    | |||  ||   |  |   |||     ||||
Db       2394 GTATTATTTTTTTTTAGTTTATTAAGGTTAGAAATTTGCGTTATTTTGTTGTTCGTTT 2453

Qy        128 TGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCT 187
              | |     ||      ||||  |   | |                | ||||||||   |           |   |
Db       2454 TTTGGTTTTATTTTTTGGATTTTATTAGGTAGTTGTTAAGTTTTATTTTGTTTGTTATTT 2513

Qy        188 GTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCT 247
               || |    ||   |   |  |      ||| ||      |   |    |  ||         |||||| |||
Db       2514 TTGTATTGTGTTTTAAATGTGTTTTTTTTAGTTTTAGTTTTTTATTATTTTTTATTGGG 2573

Qy        248 TCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATT 291
              |   ||    |   |  |   |   |||  |  |  |  |        |||  |  |
Db       2574 TTATTGTAAAGGTTTATAATTTTTTTATTTTTTTATATGTTAAT 2617

RESULT 15
US-08-913-842-6/c
; Sequence 6, Application US/08913842
; Patent No. 6028250
;  GENERAL INFORMATION:
;    APPLICANT:  OHBA, Toshiharu
;    APPLICANT:  TAKAHASHI, Shuichi
;    APPLICANT:  ANMA, Yoshiko
;    APPLICANT:  ASADA, Kiyozo
;    APPLICANT:  KATO, Ikunoshin
;    TITLE OF INVENTION:  PLANT PROMOTER AND METHOD FOR GENE
;    TITLE OF INVENTION:  EXPRESSION USING SAID PROMOTER
;    NUMBER OF SEQUENCES:  75
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK, P.L.L.C.
;      STREET:  419 7th Street N.W., Ste. 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/913,842
;      FILING DATE:
;      CLASSIFICATION:  800
```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 07-073043
; FILING DATE: 30-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00777
; FILING DATE: 26-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: OHBA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1406 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-913-842-6

Query Match 6.9%; Score 34.2; DB 3; Length 1406;
Best Local Similarity 53.3%; Pred. No. 2.5;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy        339 TTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTT 398
              | ||  |||||| |  ||  |  |  ||||   | |||      | |   | | ||| ||
Db       1118 TATTATTTATATTTTATTACAATAATATTCTAATAGATATCTAATCTCCTTTGAGACATT 1059

Qy        399 TAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTC 458
              |  | |    |||| |  ||| ||||  | |   || | |||||   || |   |||||
Db       1058 TTTTGTGTTTATTTTTATAAATTTTAAATTTCTTTATTAAAAATTATGACTCTTCTGTTG 999

Qy        459 AATTTCTTGTTAAAT 473
              | ||     |||| |
Db        998 ATTTCAAATTTAATT 984
```

Search completed: June 20, 2004, 05:03:44
Job time : 73.875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 00:55:23 ; Search time 2930.58 Seconds
(without alignments)
6182.178 Million cell updates/sec

Title: US-09-000-062-6
Perfect score: 418
Sequence: 1 tgaggtacgattcttcgatc.....taatttgttgaacagatccc 418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	418	100.0	418	6	A59349	A59349 Sequence 6
2	418	100.0	418	6	AR182671	AR182671 Sequence
3	415	99.3	4833	8	ATH3G	X60429 A.thaliana
c 4	411.8	98.5	112067	8	ATCHRIV92	AL161596 Arabidops
c 5	411.8	98.5	118267	8	ATT5J17	AL035708 Arabidops
6	54.4	13.0	5998	6	AX346536	AX346536 Sequence
7	52	12.4	349980	6	AX344561	AX344561 Sequence
8	51.2	12.2	253151	3	AE014842	AE014842 Plasmodiu
9	50.6	12.1	17527	6	AX339160	AX339160 Sequence
10	50.6	12.1	17527	6	AX346334	AX346334 Sequence
11	50.2	12.0	6306	6	AX278060	AX278060 Sequence
12	50.2	12.0	6306	6	AX323843	AX323843 Sequence
c 13	49.6	11.9	206653	2	BX470257	BX470257 Danio rer
14	49.2	11.8	127172	2	AC141780	AC141780 Apis mell
15	49.2	11.8	349980	6	AX344573	AX344573 Sequence
16	49.2	11.8	349980	6	AX344574	AX344574 Sequence
17	49	11.7	5987	6	AX346464	AX346464 Sequence
18	49	11.7	5987	6	AX458585	AX458585 Sequence
19	49	11.7	5987	6	AX795831	AX795831 Sequence
20	49	11.7	5987	6	AX795947	AX795947 Sequence
21	49	11.7	5987	6	AX822343	AX822343 Sequence
22	49	11.7	5987	6	AX822471	AX822471 Sequence
23	49	11.7	5987	6	AX825983	AX825983 Sequence
24	49	11.7	5987	6	AX826111	AX826111 Sequence
25	48.8	11.7	15479	6	AX348350	AX348350 Sequence
26	48.4	11.6	7736	3	AF310889	AF310889 Dictyoste
27	48.4	11.6	17848	6	AX277865	AX277865 Sequence
28	48.4	11.6	17848	6	AX323550	AX323550 Sequence
29	48.4	11.6	17848	6	AX348363	AX348363 Sequence
30	48	11.5	2501	6	AX598874	AX598874 Sequence
31	48	11.5	2501	6	AX599020	AX599020 Sequence
32	48	11.5	6059	6	AX346382	AX346382 Sequence
33	48	11.5	8245	6	AX251202	AX251202 Sequence
34	48	11.5	16724	6	AX281425	AX281425 Sequence
35	48	11.5	16724	6	AX345992	AX345992 Sequence
36	48	11.5	16724	6	AX348691	AX348691 Sequence
c 37	48	11.5	25568	3	AC116955	AC116955 Dictyoste
c 38	48	11.5	136791	9	AC079896	AC079896 Homo sapi
39	47.8	11.4	349980	6	AX344565	AX344565 Sequence
40	47.6	11.4	5649	6	AX767532	AX767532 Sequence
41	47.6	11.4	5649	6	AX795900	AX795900 Sequence
42	47.6	11.4	5649	6	AX822418	AX822418 Sequence
43	47.6	11.4	5649	6	AX826058	AX826058 Sequence
44	47.4	11.3	21537	6	AX346901	AX346901 Sequence
45	47.4	11.3	349980	6	AX344563	AX344563 Sequence

ALIGNMENTS

RESULT 1
A59349
LOCUS A59349 418 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 6 from Patent WO9704114.
ACCESSION A59349
VERSION A59349.1 GI:3714675
KEYWORDS .
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Derose,R., Chaubet,N. and Gigot,C.
TITLE ISOLATED DNA SEQUENCE FOR USE AS A REGULATOR REGION IN A CHIMERIC

```
            GENE USEFUL FOR TRANSFORMING PLANTS
  JOURNAL   Patent: WO 9704114-A 6 06-FEB-1997;
            RHONE POULENC AGROCHIMIE (FR)
COMMENT     Other publication FR 2736929 970124.
FEATURES              Location/Qualifiers
     source           1. .418
                      /organism="Arabidopsis thaliana"
                      /mol_type="unassigned DNA"
                      /strain="COLUMBIA"
                      /db_xref="taxon:3702"
                      /clone="C22"
                      /clone_lib="COSMID"
ORIGIN

  Query Match             100.0%;  Score 418;  DB 6;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-81;
  Matches  418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418


RESULT 2
AR182671
LOCUS       AR182671                 418 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 6 from patent US 6338961.
ACCESSION   AR182671
VERSION     AR182671.1  GI:20225878
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   DeRose,R., Chaubet,N. and Gigot,C.
  TITLE     Isolated DNA sequence capable of serving as regulatory element in a
            chimeric gene which can be used for the transformation of plants
  JOURNAL   Patent: US 6338961-A 6 15-JAN-2002;
FEATURES              Location/Qualifiers
     source           1. .418
                      /organism="unknown"
                      /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 418;  DB 6;  Length 418;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-81;
  Matches  418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418


RESULT 3
ATH3G
LOCUS       ATH3G                    4833 bp    DNA     linear   PLN 08-JAN-1996
DEFINITION  A.thaliana H3 gene 1 and H3 gene 2 for H3.3-like histone variant.
ACCESSION   X60429
VERSION     X60429.1  GI:16323
KEYWORDS    histone; histone H3; histone H3.3 homologue.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4833)
  AUTHORS   Chaubet,N., Clement,B. and Gigot,C.
  TITLE     Genes encoding a histone H3.3-like variant in Arabidopsis contain
            intervening sequences
  JOURNAL   J. Mol. Biol. 225 (2), 569-574 (1992)
  MEDLINE   92277663
   PUBMED   1593639
REFERENCE   2  (bases 1 to 4833)
  AUTHORS   Gigot,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1991) C. Gigot, Inst de Biologie Mol des Plantes,
            12 Rue du General Zimmer, 67084 Strasbourg Cedex, FRANCE
FEATURES              Location/Qualifiers
     source           1. .4833
                      /organism="Arabidopsis thaliana"
                      /mol_type="genomic DNA"
                      /strain="Columbia"
                      /db_xref="taxon:3702"
                      /cell_line="Columbia"
                      /clone_lib="cosmid c22"
                      /dev_stage="callus"
     gene             691. .1100
                      /gene="H3, gene1"
     exon             <691
                      /gene="H3, gene1"
                      /number=1
     intron           692. .1100
                      /gene="H3, gene1"
                      /number=1
     gene             join(1101. .1172,1260. .1336,1431. .1521,1620. .1790)
                      /gene="H3,gene1"
     CDS              join(1101. .1172,1260. .1336,1431. .1521,1620. .1790)
                      /gene="H3,gene1"
```

```
                /codon_start=1
                /product="histone H3.3 like protein"
                /protein_id="CAA42957.1"
                /db_xref="GI:16324"
                /db_xref="GOA:P59169"
                /db_xref="SWISS-PROT:P59169"
                /translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
                GTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAYLV
                GLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA"
     gene       1101. .1172
                /gene="H3, gene1"
     exon       1101. .1172
                /gene="H3, gene1"
                /number=2
     gene       1173. .1259
                /gene="H3, gene1"
     intron     1173. .1259
                /gene="H3, gene1"
                /number=2
     gene       1260. .1336
                /gene="H3, gene1"
     exon       1260. .1336
                /gene="H3, gene1"
                /number=3
     gene       1337. .1430
                /gene="H3, gene1"
     intron     1337. .1430
                /gene="H3, gene1"
                /number=3
     gene       1431. .1521
                /gene="H3, gene1"
     exon       1431. .1521
                /gene="H3, gene1"
                /number=4
     gene       1522. .1619
                /gene="H3, gene1"
     intron     1522. .1619
                /gene="H3, gene1"
                /number=4
     gene       1620. .1790
                /gene="H3, gene1"
     exon       1620. .>1790
                /gene="H3, gene1"
                /number=5
     gene       join(2912. .2983,3065. .3141,3227. .3488)
                /gene="H3,gene 2"
     CDS        join(2912. .2983,3065. .3141,3227. .3488)
                /gene="H3,gene 2"
                /codon_start=1
                /product="histone H3.3 like protein"
                /protein_id="CAA42958.1"
                /db_xref="GI:404825"
                /db_xref="GOA:P59169"
                /db_xref="SWISS-PROT:P59169"
                /translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
                GTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAYLV
                GLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA"
     gene       2912. .2983
                /gene="H3, gene 2"
     exon       <2912. .2983
                /gene="H3, gene 2"
                /number=1
     gene       2984. .3064
                /gene="H3, gene 2"
     intron     2984. .3064
                /gene="H3, gene 2"
                /number=1
     gene       3065. .3141
                /gene="H3, gene 2"
     exon       3065. .3141
                /gene="H3, gene 2"
                /number=2
     gene       3142. .3226
                /gene="H3, gene 2"
     intron     3142. .3226
                /gene="H3, gene 2"
                /number=2
     gene       3227. .3488
                /gene="H3, gene 2"
     exon       3227. .>3488
                /gene="H3, gene 2"
                /number=3
ORIGIN

  Query Match           99.3%;  Score 415;  DB 8;  Length 4833;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-81;
  Matches  415;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        688 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 747

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        748 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 807

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        808 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 867

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        868 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 927

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        928 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 987

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        988 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 1047

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGAT 415
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1048 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGAT 1102


RESULT 4
ATCHRIV92/c
LOCUS       ATCHRIV92              112067 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 92.
ACCESSION   AL161596
VERSION     AL161596.2  GI:7271037
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 107700)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 9546 to 9977)
  AUTHORS   Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 107578 to 112067)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 112067)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
```

```
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV91 at the 5' end and an
            overlap with ATCHRIV93 at the 3' end.
FEATURES            Location/Qualifiers
     source         1. .112067
                    /organism="Arabidopsis thaliana"
                    /mol_type="genomic DNA"
                    /variety="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="4"
     gene           6122. .8203
                    /gene="AT4g39830"
     CDS            join(6122. .6415,6481. .6573,6652. .6864,6959. .7153,
                    7250. .8203)
                    /gene="AT4g39830"
                    /note="strong similarity to L-ascorbate oxidase, Cucumis
                    sativus, PIR1:KSKVAO
                    Contains Glycosyl hydrolases family 5 signature
                    AA384-393;Multicopper oxidases signatures
                    AA131-151;Multicopper oxidases signatures
                    AA540-560;Multicopper oxidases signatures
                    AA545-556;Prokaryotic membrane lipoprotein lipid
                    attachment site AA567-577
                    contains EST gb:AI994215.1, AA585817, AI996008.1"
                    /codon_start=1
                    /product="putative L-ascorbate oxidase"
                    /protein_id="CAB80646.1"
                    /db_xref="GI:7271038"
                    /db_xref="GOA:O65670"
                    /db_xref="SPTREMBL:O65670"
                    /translation="MMRPKRSSDTVHVFNLMVLCFIALFFSSVLCQGKIRRFKWEVKY
                    EFKSPDCFEKLVITINGKFPGPTIKAQQGDTIVVELKNSFMTENVAVHWHGIRQIGTP
                    WFDGVEGVTQCPILPGEVFIYQFVVDRPGTYMYHSHYGMQRESGLIGMIQVSPPATEP
                    EPFTYDYDRNFLLTDWYHKSMSEKATGLASIPFKWVGEPQSLMIQGRGRFNCSNNLTT
                    PPSLVSGVCNVSNADCSRFILTVIPGKTYRLRIGSLTALSALSFQIEGHNLTVVEADG
                    HYVEPFTVKNLFVYSGETYSVLLKADQNPRRNYWITSSIVSRPATTPPATAVLNYYPN
                    HPRRRPPTSESSNIVPEWNDTRSRLAQSLAIKARRGFIHALPENSDKVIVLLNTQNEV
                    NGYRRWSVNNVSYHHPKTPYLIALKQNLTNAFDWRFTAPENYDSRNYDIFAKPLNANA
                    TTSDGIYRLRFNSTVDVILQNANTMNANNSETHPWHLHGHDFWVLGYGEGKFNESEDP
                    KRYNRVDPIKKNTVAVQPFGWTALRFRADNPGVWSFHCHIESHFFMGMGIVFESGIDK
                    VSSLPSSIMGCGQTKR"
     exon           6122. .6415
                    /gene="AT4g39830"
                    /number=1
     intron         6416. .6480
                    /gene="AT4g39830"
                    /number=1
     exon           6481. .6573
                    /gene="AT4g39830"
                    /number=2
     intron         6574. .6651
                    /gene="AT4g39830"
                    /number=2
     exon           6652. .6864
                    /gene="AT4g39830"
                    /number=3
     intron         6865. .6958
                    /gene="AT4g39830"
                    /number=3
     exon           6959. .7153
                    /gene="AT4g39830"
                    /number=4
     intron         7154. .7249
                    /gene="AT4g39830"
                    /number=4
     exon           7250. .8203
                    /gene="AT4g39830"
                    /number=5
     gene           12287. .13642
                    /gene="AT4g39840"
     CDS            12287. .13642
                    /gene="AT4g39840"
                    /note="contains EST gb:AI999725.1, Z46537, F13826,
                    AI996686.1"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB80647.1"
                    /db_xref="GI:7271039"
                    /db_xref="SPTREMBL:Q9SMR9"
                    /translation="MSPLKYLCFRKVLLLLLYSSLFLFSSSSFFASSVDDHELQRIDP
                    HFRVRRLLVKDLDSTGDDDETNLPPPPKKKKLLTDSVSSSSSSSSGTKKNQTKLLKPIS
                    SSSSTKNQTKLAKTTTMGTSHKLNSTKSSSNTTKTSSELKKLNSGTKSTNSTSSIKKS
                    ADLSKSSSSKNKTTIKPPSSKLSSPPSEKKSQPSSKPVTKSKQSEKEIKPFWLDDEED
                    EDFVSEFRDLPTRFQRSLIPDLERISTTSKNYINKANKQITKNFKPYFGNKYAPTIAS
                    VVSFVFILVPLLLVSLIFNRFKAYFSLQKILIFIQIYLSIYFSILCLSSLVTGIEPLK
                    FLYATSSSTYVCLQILQTLGYVFYLLLLLMYLVLVFSTDCGLGLKVLGLAQTFVGFAV
                    GLHYYVAVFHRVVLRQPPKTNWKIHGVYATCFLLICLLSNAERRKKEYLEEGGDEGKK
                    N"
     exon           12287. .13642
                    /gene="AT4g39840"
                    /number=1
     misc_feature   15539. .16205
                    /note="EST N95877 matches to positions 26106-26355,
                    26723-26772; possibly related to subsequent gen"
     gene           16239. .23312
                    /gene="AT4g39850"
     CDS            join(16239. .16300,16394. .16679,16981. .17205,17341. .17538,
                    17675. .18129,18222. .18357,18433. .18511,19039. .19271,
                    19326. .19630,19712. .19844,19901. .20056,20213. .20435,
                    20515. .20623,20845. .20935,21205. .21372,21453. .21568,
                    21626. .21860,21924. .22104,22162. .22362,22534. .22680,
                    22750. .22983,23134. .23312)
                    /gene="AT4g39850"
                    /note="similarity to 70K peroxisomal membrane protein,
                    Rattus norvegicus , PIR2:A35723
                    Contains ABC transporters family signature AA580-594;ABC
                    transporters family signature AA1369-1383;ATP/GTP-binding
                    site motif A (P-loop) AA1247-1254
                    contains EST gb:N95877"
                    /codon_start=1
                    /product="putative protein"
                    /protein_id="CAB80648.1"
                    /db_xref="GI:7271040"
                    /db_xref="GOA:Q9SMR8"
                    /db_xref="SPTREMBL:Q9SMR8"
                    /translation="MPSLQLLQLTERGRGLVASRRKSILLAAGIVAAGGTAVYLKSRV
                    ASRRPDSSRLCNGQSDDDETLEKLTATDQNAKITTKKKKGGGLKSLQVLTAILLSQMG
                    KMGARDLLALVATVVFRTALSNRLAKVQGFLFRAAFLRRAPLFLRLISENIMLCFMLS
                    TLHSTSKYITGALSLRFRKILTKIIHSHYFENMVYYKISHVDGRITHPEQRIASDVPR
                    FSSELSDLILDDLTAVTDGILYAWRLCSYASPKYIFWILAYVLGAGTAIRNFSPSFGK
                    LMSKEQQLEGEYRQLHSRLRTHSESIAFYGGETREESHIQQKFKNLVSHMSHVLHDHW
                    WFGMIQDFLLKYLGATVAVILIIEPFFSGHLRPDDSTLGRAEMLSNIRYHTSVIISLF
                    QALGTLSISSRRLNRLSGYADRIHELMAVSRELSGDDKSSFQRNRSRNYLSEANYVEF
                    SDVKVVTPTGNVLVEDLTLRVEQGSNLLITGGLWPLVSGHIVKPGVGSDLNKEIFYVP
                    QRPYMAVGTLRDQLIYPLTSGQESELLTEIGMVELLKNVSILVSLFYTRMVDLEYLLD
                    RYQPEKEVNWGDELSLGEQQRLGMARLFYHKPKFAILDECTSAVTTDMEERFAAKVRA
                    MGTSCITISHRPALVAFHDVVLSLDGEGGWSVHYKRDDSALLTDAEIDSVKSSDTDRQ
                    NDAMVVQRAFAAARKVSAEVLFESATNSKAQSYQTQLIARSPVVDKSVVLPRFPQPQT
                    SQRALPSRVAAMLNVLLPWKPHLLFCISTWLTYSSCADANIDIILLKITDHFIVTFQI
                    PTIFDKQGAQLLAVACLVVSRTLISDRIASLNGTTVKYVLEQDKAAFVRLIGLSVLQS
                    GASSIIAPSLRHLTQRLALGWRIRLTQHLLRNYLRNNAFYKVFHMSGNSIDADQRLTR
                    DLEKLTADLSGLLTGMVKPSVDILWYVRYVGLISNICFMRGVAILYTYMLLGLGFLRR
                    VAPDFGDLAGEEQQLEGKFRPTSSCLCRFMHERLNTHAESIAFFGGGAREKAVSFLIA
                    LAIAAGFWVVIFESKLPEVNLELFYHLCDDGLFMSNVASLMVDKKFRALLDHSLMLLR
                    KKWLYGILDDFVTKQLPNNVTWGLSLLYALEHKGDRALVSTQVKYLHVLYMMLLIPGE
                    LAHALRYLASVVSQSFMAFGDILELHKKFLELSGGINRIFELDEFLDASQSGVTSENQ
                    TSRLDSQDLLSFSEVDIITPAQKLMASKLSCEIVSGKSLLVTGPNGSGKTSVFRVLRD
                    IWPTVCGRLTKPSLDIKELGSGNGMFFVPQRPYTCLGTLRDQIIYPLSKEEAEKRAAK
                    LYTSGESSTEAGSILDSHLKTILENVRLVYLLERDVGGWDATTNWEDILSLGEQQRLG
                    MVSLL"
```

```
     exon            16239. .16300
                     /gene="AT4g39850"
                     /number=1
     intron          16301. .16393
                     /gene="AT4g39850"
                     /number=1
     exon            16394. .16679
                     /gene="AT4g39850"
                     /number=2
     intron          16680. .16980
                     /gene="AT4g39850"
                     /number=2
     exon            16981. .17205
                     /gene="AT4g39850"
                     /number=3
     intron          17206. .17340
                     /gene="AT4g39850"
                     /number=3
     exon            17341. .17538
                     /gene="AT4g39850"
                     /number=4
     intron          17539. .17674
                     /gene="AT4g39850"
                     /number=4
     exon            17675. .18129
                     /gene="AT4g39850"
                     /number=5

  Query Match             98.5%;  Score 411.8;  DB 8;  Length 112067;
  Best Local Similarity   99.5%;  Pred. No. 2.4e-80;
  Matches  413;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85668 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 85609

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85608 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 85549

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      85548 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGGATAAAGATTTTAGGTTAGAGATGAA 85489

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85488 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 85429

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85428 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 85369

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85368 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 85309

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGAT 415
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85308 CATTCCGGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGAT 85254


RESULT 5
ATT5J17/c
LOCUS       ATT5J17               118267 bp    DNA     linear   PLN 21-SEP-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T5J17 (ESSA
            project).
ACCESSION   AL035708
VERSION     AL035708.2  GI:5918309
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Murphy,G., Ridley,P., Hudson,S., Bancroft,I., Mewes,H.W.,
            Mayer,K.F.X., Lemcke,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118267)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Sep 22, 1999 this sequence version replaced gi:4490734.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .118267
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     misc_feature    1. .22789
                     /note="position 1-22789 overlaps to BAC T19P19, EMBL
                     accession AL022605; please refer to this entry for
                     analysis and annotation"
     gene            22854. .24209
                     /gene="T5J17.10"
     CDS             22854. .24209
                     /gene="T5J17.10"
                     /note="contains EST gb:F13826, Z46537"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB38897.1"
                     /db_xref="GI:4490735"
                     /db_xref="SPTREMBL:Q9SMR9"
                     /translation="MSPLKYLCFRKVLLLLLYSSLFLFSSSSFFASSVDDHELQRIDP
                     HFRVRRLLVKDLDSTGDDDETNLPPPPKKKKLLTDSVSSSSSSSGTKKNQTKLLKPIS
                     SSSSTKNQTKLAKTTTMGTSHKLNSTKSSSNTTKTSSELKKLNSGTKSTNSTSSIKKS
                     ADLSKSSSSKNKTTIKPPSSKLSSPPSEKKSQPSSKPVTKSKQSEKEIKPFWLDDEED
                     EDFVSEFRDLPTRFQRSLIPDLERISTTSKNYINKANKQITKNFKPYFGNKYAPTIAS
                     VVSFVPILVPLLLVSLIFNRFKAYFSLQKILIFIQIYLSIYFSILCLSSLVTGIEPLK
                     FLYATSSSTYVCLQILQTLGYVFYLLLLLMYLVLVFSTDCGLGLKVLGLAQTFVGFAV
                     GLHYYVAVFHRVVLRQPPKTNWKIHGVYATCFLLICLLSNAERRKKEYLEEGGDEGKK
                     N"
     exon            22854. .24209
                     /gene="T5J17.10"
                     /number=1
     misc_feature    26106. .26772
                     /note="EST N95877 matches to positions 26106-26355,
                     26723-26772; possibly related to subsequent gen"
     gene            26806. .33879
                     /gene="T5J17.20"
     CDS             join(26806. .26867,26961. .27246,27548. .27772,27908. .28105,
                     28242. .28696,28789. .28924,29000. .29078,29606. .29838,
                     29893. .30197,30279. .30411,30468. .30623,30780. .31002,
                     31082. .31190,31412. .31502,31772. .31939,32020. .32135,
                     32193. .32427,32491. .32671,32729. .32929,33101. .33247,
                     33317. .33550,33701. .33879)
                     /gene="T5J17.20"
                     /note="similarity to 70K peroxisomal membrane protein,
                     Rattus norvegicus , PIR2:A35723
                     Contains ABC transporters family signature
                     [LSLGEQQRLGMARLF][LSLGEQQRLGMVSLL]"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB38898.1"
                     /db_xref="GI:4490736"
                     /db_xref="GOA:Q9SMR8"
                     /db_xref="SPTREMBL:Q9SMR8"
```

/translation="MPSLQLLQLTERGRGLVASRRKSILLAAGIVAAGGTAVYLKSRV
ASRRPDSSRLCNGQSDDDETLEKLTATDQNAKITTKKKKGGGLKSLQVLTAILLSQMG
KMGARDLLALVATVVFRTALSNRLAKVQGFLFRAAFLRRAPLFLRLISENIMLCFMLS
TLHSTSKYITGALSLRFRKILTKIIHSHYFENMVYYKISHVDGRITHPEQRIASDVPR
FSSELSDLILDDLTAVTDGILYAWRLCSYASPKYIFWILAYVLGAGTAIRNFSPSFGK
LMSKEQQLEGEYRQLHSRLRTHSESIAFYGGETREESHIQQKFKNLVSHMSHVLHDHW
WFGMIQDFLLKYLGATVAVILIIEPPFFSGHLRPDDSTLGRAEMLSNIRYHTSVIISLF
QALGTLSISSSRRLNRLSGYADRIHELMAVSRELSGDDKSSFQRNRSRNYLSEANYVBF
SDVKVVTPTGNVLVEDLTLRVEQGSNLLITGGLWPLVSGHIVKPGVGSDLNKEIFYVP
QRPYMAVGTLRDQLIYPLTSGQESELLTEIGMVELLKNVSILVSLFYTRMVDLEYLLD
RYQPEKEVNWGDELSLGEQQRLGMARLFYHKPKFAILDECTSAVTTDMEERFAAKVRA
MGTSCITISHRPALVAFHDVVLSLDGEGGWSVHYKRDDSALLTDAEIDSVKSSDTDRQ
NDAMVVQRAFAAARKVSAEVLFESATNSKAQSYQTQLIARSPVVDKSVVLPRFPQPQT
SQRALPSRVAAMLNVLLPWKPHLLFCISTWLTYSSCADANIDIILLKITDHFIVTFQI
PTIFDKQGAQLLAVACLVVSRTLISDRIASLNGTTVKYVLEQDKAAFVRLIGLSVLQS
GASSIIAPSLRHLTQRLALGWRIRLTQHLLRNYLRNNAFYKVFHMSGNSIDADQRLTR
DLEKLTADLSGLLTGMVKPSVDILWYVRYVGLISNICFMRGVAILYTYMLLGLGFLRR
VAPDFGDLAGEEQQLEGKFRPTSSCLCRFMHERLNTHAESIAFFGGGAREKAVSFLIA
LAIAAGFWVVIFESKLPEVNLELFYHLCDDGLFMSNVASLMVDKKFRALLDHSLMLLR
KKWLYGILDDFVTKQLPNNVTWGLSLLYALEHKGDRALVSTQVKYLHVLYMMLLIPGE
LAHALRYLASVVSQSFMAFGDILELHKKFLELSGGINRIFELDEFLDASQSGVTSENQ
TSRLDSQDLLSFSEVDIITPAQKLMASKLSCEIVSGKSLLVTGPNGSGKTSVFRVLRD
IWPTVCGRLTKPSLDIKELGSGNGMFFVPQRPYTCLGTLRDQIIYPLSKEEAEKRAAK
LYTSGESSTEAGSILDSHLKTILENVRLVYLLERDVGGWDATTNWEDILSLGEQQRLG
MVSLL"

exon	26806. .26867 /gene="T5J17.20" /number=1
intron	26868. .26960 /gene="T5J17.20" /number=1
exon	26961. .27246 /gene="T5J17.20" /number=2
intron	27247. .27547 /gene="T5J17.20" /number=2
exon	27548. .27772 /gene="T5J17.20" /number=3
intron	27773. .27907 /gene="T5J17.20" /number=3
exon	27908. .28105 /gene="T5J17.20" /number=4
intron	28106. .28241 /gene="T5J17.20" /number=4
exon	28242. .28696 /gene="T5J17.20" /number=5
intron	28697. .28788 /gene="T5J17.20" /number=5
exon	28789. .28924 /gene="T5J17.20" /number=6
intron	28925. .28999 /gene="T5J17.20" /number=6
exon	29000. .29078 /gene="T5J17.20" /number=7
intron	29079. .29605 /gene="T5J17.20" /number=7
exon	29606. .29838 /gene="T5J17.20" /number=8
intron	29839. .29892 /gene="T5J17.20" /number=8
exon	29893. .30197 /gene="T5J17.20" /number=9
intron	30198. .30278 /gene="T5J17.20" /number=9
exon	30279. .30411 /gene="T5J17.20" /number=10
intron	30412. .30467 /gene="T5J17.20" /number=10
exon	30468. .30623 /gene="T5J17.20" /number=11
intron	30624. .30779 /gene="T5J17.20" /number=11
exon	30780. .31002 /gene="T5J17.20" /number=12
intron	31003. .31081 /gene="T5J17.20" /number=12
exon	31082. .31190 /gene="T5J17.20" /number=13
intron	31191. .31411 /gene="T5J17.20" /number=13
exon	31412. .31502 /gene="T5J17.20" /number=14
intron	31503. .31771 /gene="T5J17.20" /number=14
exon	31772. .31939 /gene="T5J17.20" /number=15
intron	31940. .32019 /gene="T5J17.20" /number=15
exon	32020. .32135 /gene="T5J17.20"

Query Match 98.5%; Score 411.8; DB 8; Length 118267;
Best Local Similarity 99.5%; Pred. No. 2.4e-80;
Matches 413; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96235 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 96176

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96175 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 96116

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      96115 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGGATAAAGATTTTAGGTTAGAGATGAA 96056

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96055 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 95996

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95995 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 95936

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      95935 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 95876
```

```
Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGAT 415
Db      95875 CATTCCGGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGAT 95821
```

RESULT 6
AX346536

```
LOCUS       AX346536                5998 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1607 from Patent WO0200928.
ACCESSION   AX346536
VERSION     AX346536.1  GI:18494422
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1607 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1..5998
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             13.0%;  Score 54.4;  DB 6;  Length 5998;
  Best Local Similarity   49.7%;  Pred. No. 0.045;
  Matches  165;  Conservative    0;  Mismatches  166;  Indels    1;  Gaps    1;

Qy         78 GATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTT 137
Db       4382 GAGTAGTGGTTTGAGTATAGTTAATTGTAGTTTCGAATTTTGGGTTTAAATGATTTTTT 4441

Qy        138 CGCCGAATGATCTTCCGTA-TAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTT 196
Db       4442 TGTTTTAGTTTTTTGATTAGTTGGGATTATAGGTTATAGTTATGTTGTTTAGTTAATTAA 4501

Qy        197 CATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGT 256
Db       4502 AAAAAAAAATTGTTTTTTTTTTTATAGAGATAGAAGTTTTTTATGTTGTTTAGGTTGG 4561

Qy        257 AAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGA 316
Db       4562 TTTTGAATTTTTGGTTTTAGGCGATTTTTTATTTTTTTTTAGTTTTTGTGTTATTATA 4621

Qy        317 TTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAA 376
Db       4622 TTTTTAGGGTAATTTTTTATTTGTTATTTATTATTTTTTGTATTTTTTTTTTTAGGTTTT 4681

Qy        377 TACGTATGGATCTAAATCTTGTTAATTTGTTG 408
Db       4682 TATGGGATTATTTTTTGTTTTTTATGTTTTAG 4713
```

RESULT 7
AX344561

```
LOCUS       AX344561              349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 12 from Patent WO0200932.
ACCESSION   AX344561
VERSION     AX344561.1  GI:18492447
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of known genetic parameters within the mhc
  JOURNAL   Patent: WO 0200932-A 12 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1..349980
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo
                     sapiens)~Original length of seq 1: 3.673778 <223>~split as
                     follows.~seq 01 0.000.001 TO 0.349.980~seq 02 0.300.001
                     649.980~seq 03 600.001 949.980~seq 04 900.001
                     1.249.980~seq 05 1.200.001 1.549.980~seq 06 1.500.001
                     1.849.980~seq 07 1.800.001 2.149.980~seq 08 2.100.001
                     2.449.980~seq 09 2.400.001 2.749.980~seq 10 2.700.001
                     3.049.980~seq 11 3.000.001 3.349.980~seq 12 3.300.001
                     3.649.980~seq 13 3.600.001 3.673.778 <223>~Original length
                     of seq 2: 3.673778 <223>~split as follows.~seq 14
                     0.000.001 TO 0.349.980~seq 15 0.300.001 649.980~seq 16
                     600.001 949.980~seq 17 900.001 1.249.980~seq 18 1.200.001
                     1.549.980~seq 19 1.500.001 1.849.980~seq 20 1.800.001
                     2.149.980~seq 21 2.100.001 2.449.980~seq 22 2.400.001
                     2.749.980~seq 23 2.700.001 3.049.980~seq 24 3.000.001
                     3.349.980~seq 25 3.300.001 3.649.980~seq 26 3.600.001
                     3.673.778"
ORIGIN

  Query Match             12.4%;  Score 52;  DB 6;  Length 349980;
  Best Local Similarity   47.8%;  Pred. No. 0.091;
  Matches  151;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy         38 AAATATTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGTGGTACGAAATTAG 97
Db     277723 AATTATTATTTTTGTGGTCGTTTTTTTATGTATTTTTCGGTAAAGATTTATTGGAATTTT 277782

Qy         98 GCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTAT 157
Db     277783 TTTAGGTAATTTTTTATTTTTTCGATGTTTTTTTATTTTTTATTTTCGAAGTTTAATTAT 277842

Qy        158 AAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTG 217
Db     277843 ATATATTATTGAAATAATTGGAGTTTTTTAGTTTTGTATTATAGATAAGTTTTTTATTAA 277902

Qy        218 TCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGG 277
Db     277903 TATTGGTAAATTTAAATATATTAGTAGTTATTATTGTTTAGGAAGATATTTATTTTGTAT 277962

Qy        278 CGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTT 337
Db     277963 TTTATTTTATTTTTGGTTATTTTTTTTTTGTTTGTTAGATTTTTTTTTTAGGTTTTTTTT 278022

Qy        338 GGTTTTGTGTTGATCT 353
Db     278023 TTGTTTGTTTATATTT 278038
```

RESULT 8
AE014842

```
LOCUS       AE014842              253151 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 11 section 7 of 8 of the
            complete sequence.
ACCESSION   AE014842 AE014186
VERSION     AE014842.1  GI:23496321
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253151)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
```

```
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253151)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253151
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="11"
     repeat_region   complement(73. .101)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   173. .279
                     /rpt_type=tandem
                     /rpt_unit="(TAAAAA)n"
     repeat_region   complement(302. .328)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(393. .466)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   393. .442
                     /rpt_type=tandem
     repeat_region   440. .466
                     /rpt_type=tandem
     repeat_region   545. .578
                     /rpt_type=tandem
     repeat_region   complement(545. .578)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(580. .673)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   762. .795
                     /rpt_type=tandem
     repeat_region   complement(762. .795)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(869. .955)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(956. .975)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(983. .1040)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1094. .1121
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(1149. .1221)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1239. .1281
                     /rpt_type=tandem
     repeat_region   complement(1243. .1319)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(1323. .1379)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1571. .1605)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1621. .1699)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1820. .1872)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1921. .2010)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2105. .2135
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(2135. .2176)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(2226. .2248)
                     /rpt_type=tandem
                     /rpt_unit="(TAAAA)n"
     repeat_region   complement(2289. .2362)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2372. .2396
                     /rpt_type=tandem
     repeat_region   complement(2372. .2396)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(2402. .2424)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2438. .2620)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2636. .2671
                     /rpt_type=tandem
     repeat_region   complement(2636. .2671)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   2680. .2837
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
     repeat_region   complement(2763. .2865)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(2932. .2954)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2970. .3008
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(3028. .3145)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3189. .3238)
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
     repeat_region   complement(3250. .3403)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   3425. .3520
                     /rpt_type=tandem
     repeat_region   complement(3425. .3520)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   3556. .3580
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(3584. .3641)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3649. .3809)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(3842. .3884)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
```

```
     repeat_region   3845. .3884
                     /rpt_type=tandem
     repeat_region   3897. .3944
                     /rpt_type=tandem
                     /rpt_unit="(TATAA)n"
     repeat_region   complement(3946. .3999)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   4000. .4024
                     /rpt_type=tandem
     repeat_region   complement(4000. .4024)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(4026. .4104)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   4163. .4226
                     /rpt_type=tandem
                     /rpt_unit="(TAAAA)n"
     repeat_region   4232. .4267
                     /rpt_type=tandem
     repeat_region   complement(4232. .4267)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   complement(4339. .4359)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
     repeat_region   4380. .4407
                     /rpt_type=tandem
     repeat_region   complement(4380. .4407)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
     repeat_region   complement(4408. .4575)
                     /note="AT_rich"
                     /rpt_type=tandem

  Query Match             12.2%;  Score 51.2;  DB 3;  Length 253151;
  Best Local Similarity   49.6%;  Pred. No. 0.14;
  Matches  131;  Conservative    0;  Mismatches  133;  Indels    0;  Gaps    0;

Qy        155 TATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCT 214
              ||| |  || ||    || | |   | | ||||     || |   |  | ||  ||| |
Db       3286 TATTATTATATTTTTATATATACATAACATATATTATTATATATATAATATATTTTTTGT 3345

Qy        215 TTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATT 274
              | |    ||| | |   ||     ||| | ||   || | |   || ||  || |  ||
Db       3346 ATATAATTAATTTATATTATAAAATATAATATTTTTATTTATGTTTATATATTTATATTT 3405

Qy        275 CGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTC 334
              |   ||| ||  |||   ||||  |||| || |||||   |||    ||    ||||
Db       3406 TGATTATTATTTGTTATTATTTTTTTTTTTTTTATTATTTTATTATTTTTTTTTTTTTT 3465

Qy        335 TTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATC 394
              | |  |||| | |   | |||   |   |  |   |||   ||  | ||  | |||
Db       3466 TATTATTTTTTTTATTTTTTATTTTTTTATTTTTTTATTTTTATTTTTGTTTTTAAGCA 3525

Qy        395 TTGTTAATTTGTTGAACAGATCCC 418
              ||   ||| |  |||| |  || |
Db       3526 TTAAAAATATCATGAATATTTCAC 3549


RESULT 9
AX339160
LOCUS       AX339160               17527 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 27 from Patent WO0176451.
ACCESSION   AX339160
VERSION     AX339160.1  GI:18129237
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with metabolism
  JOURNAL   Patent: WO 0176451-A 27 18-OCT-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .17527
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             12.1%;  Score 50.6;  DB 6;  Length 17527;
  Best Local Similarity   52.1%;  Pred. No. 0.27;
  Matches  113;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        192 GATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACA 251
              | | | || |  | ||  | | |||   | || ||  |   ||| | || ||||||  | |
Db      16590 GTTGTTATAAGAATATTTTGTTTTTTATTTGAGTCGTTATTATTTTAGAAAGTTTTAAGA 16649

Qy        252 TGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTA 311
              |||  | | | | ||| || || |  | || ||  ||    ||||  |||| ||  ||
Db      16650 GGTGGTATTTGTTGGTTGTTTTTTGTTGGTTTTTTTTTTTTTATTTTTTTTTTTTTAGTTT 16709

Qy        312 TGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAA 371
              |    | ||    | ||||  || |||  || | | |   | ||    |  |||  | |
Db      16710 TAATAGTGATTATTGGGAAGTTTTTTTTTTGTTTTTGTTTGTTTTTTTGTATTTTTTTAT 16769

Qy        372 TTGAATACGTATGGATCTAAATCTTGTTAATTTGTTG 408
              ||  |  | |  | |   | || ||  ||| | |
Db      16770 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAG 16806


RESULT 10
AX346334
LOCUS       AX346334               17527 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1405 from Patent WO0200928.
ACCESSION   AX346334
VERSION     AX346334.1  GI:18494220
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1405 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .17527
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             12.1%;  Score 50.6;  DB 6;  Length 17527;
  Best Local Similarity   52.1%;  Pred. No. 0.27;
  Matches  113;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy        192 GATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACA 251
              | | | || |  | ||  | | |||   | || ||  |   ||| | || ||||||  | |
Db      16590 GTTGTTATAAGAATATTTTGTTTTTTATTTGAGTCGTTATTATTTTAGAAAGTTTTAAGA 16649

Qy        252 TGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTA 311
              |||  | | | | ||| || || |  | || ||  ||    ||||  |||| ||  ||
Db      16650 GGTGGTATTTGTTGGTTGTTTTTTGTTGGTTTTTTTTTTTTTATTTTTTTTTTTTTAGTTT 16709

Qy        312 TGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAA 371
              |    | ||    | ||||  || |||  || | | |   | ||    |  |||  | |
Db      16710 TAATAGTGATTATTGGGAAGTTTTTTTTTTGTTTTTGTTTGTTTTTTTGTATTTTTTTAT 16769

Qy        372 TTGAATACGTATGGATCTAAATCTTGTTAATTTGTTG 408
              ||  |  | |  | |   | || ||  ||| | |
```

```
Db      16770 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAG 16806


RESULT 11
AX278060
LOCUS       AX278060                6306 bp    DNA     linear   PAT 01-NOV-2001
DEFINITION  Sequence 223 from Patent WO0177375.
ACCESSION   AX278060
VERSION     AX278060.1  GI:16605132
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with gene regulation
  JOURNAL   Patent: WO 0177375-A 223 18-OCT-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match           12.0%;  Score 50.2;  DB 6;  Length 6306;
  Best Local Similarity   50.2%;  Pred. No. 0.37;
  Matches  124;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        161 GATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCT 220
              || ||| |||| |||  | || |||| | |||||||  | |  |  | ||||  |  ||
Db       5490 GAGTTTCGGTTCGAGGGGTATAGTATGGTTAGATTTAGTGATTAATTTGTTTTTAGGTAG 5549

Qy        221 AGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGA 280
              |   | ||  | |  |    ||||     | | ||   |    |  || || ||
Db       5550 ATGTTTTCGGTATTGATTTTAGATTTTTTGTTTTTTAGTTTATTGTTTTTTTTTTAAGTT 5609

Qy        281 TTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGT 340
              |  || |  |    ||||| |    || ||    || ||    ||  |  |||| |||  |
Db       5610 TGATTTAATTTATTTTTGGTGGGTTTTATATATCGTTTTTTTTTTTGTTTTTTTTTTTTT 5669

Qy        341 TTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTA 400
              ||| | || || ||    |  ||  ||   ||   |  | |   | |   | || ||
Db       5670 TTTTTTTTTATTTTTTTTTTTTTTTTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5729

Qy        401 ATTTGTT 407
              ||| ||
Db       5730 TTTTTTT 5736


RESULT 12
AX323843
LOCUS       AX323843                6306 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 331 from Patent WO0192565.
ACCESSION   AX323843
VERSION     AX323843.1  GI:18094592
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with dna transcription
  JOURNAL   Patent: WO 0192565-A 331 06-DEC-2001;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6306
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match           12.0%;  Score 50.2;  DB 6;  Length 6306;
  Best Local Similarity   50.2%;  Pred. No. 0.37;
  Matches  124;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        161 GATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCT 220
              || ||| |||| |||  | || |||| | |||||||  | |  |  | ||||  |  ||
Db       5490 GAGTTTCGGTTCGAGGGGTATAGTATGGTTAGATTTAGTGATTAATTTGTTTTTAGGTAG 5549

Qy        221 AGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGA 280
              |   | ||  | |  |    ||||     | | ||   |    |  || || ||
Db       5550 ATGTTTTCGGTATTGATTTTAGATTTTTTGTTTTTTAGTTTATTGTTTTTTTTTTAAGTT 5609

Qy        281 TTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGT 340
              |  || |  |    ||||| |    || ||    || ||    ||  |  |||| |||  |
Db       5610 TGATTTAATTTATTTTTGGTGGGTTTTATATATCGTTTTTTTTTTTGTTTTTTTTTTTTT 5669

Qy        341 TTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTA 400
              ||| | || || ||    |  ||  ||   ||   |  | |   | |   | || ||
Db       5670 TTTTTTTTTATTTTTTTTTTTTTTTTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5729

Qy        401 ATTTGTT 407
              ||| ||
Db       5730 TTTTTTT 5736


RESULT 13
BX470257/c
LOCUS       BX470257              206653 bp    DNA     linear   HTG 24-SEP-2003
DEFINITION  Danio rerio clone DKEY-99O15, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   BX470257
VERSION     BX470257.3  GI:35209325
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 206653)
  AUTHORS   Mclaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2003 this sequence version replaced gi:32169046.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK99O15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 205905 bases at least Q40
            Consensus quality: 206063 bases at least Q30
            Consensus quality: 206151 bases at least Q20
            Insert size: 206253; sum-of-contigs
            Insert size: 198710; 1.2% error; agarose-fp
            Quality coverage: 7.80x in Q20 bases; sum-of-contigs Quality
            coverage: 8.14x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64965: contig of 64965 bp in length
```

```
            *    64966    65065: gap of 100 bp
            *    65066    83771: contig of 18706 bp in length
            *    83772    83871: gap of 100 bp
            *    83872   136962: contig of 53091 bp in length
            *   136963   137062: gap of 100 bp
            *   137063   186755: contig of 49693 bp in length
            *   186756   186855: gap of 100 bp
            *   186856   206653: contig of 19798 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206653
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-99O15"
                     /clone_lib="DanioKey"
     misc_feature    1. .64965
                     /note="assembly_fragment:03254
                     fragment_chain:1"
     misc_feature    65066. .83771
                     /note="assembly_fragment:01800
                     fragment_chain:1"
     misc_feature    83872. .136962
                     /note="assembly_fragment:02854
                     fragment_chain:1"
     misc_feature    137063. .186755
                     /note="assembly_fragment:00600.0"
     misc_feature    186856. .206653
                     /note="assembly_fragment:01616.0"
ORIGIN

  Query Match             11.9%;  Score 49.6;  DB 2;  Length 206653;
  Best Local Similarity   47.7%;  Pred. No. 0.33;
  Matches  145;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        104 TTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGAT 163
              ||   || |||| |     ||| | || |||| |      || ||  |   | ||     |
Db      33798 TTTTATTTTATTTTGTTTTATTTTATTATTTCATTTTATTATTATTGTTTTTTTATTATT 33739

Qy        164 TTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGA 223
              ||   ||| |  |  || | ||      |||| |     ||| ||| ||    || |
Db      33738 ATTTTATTATATTTTTATTGCATTTTATAATTTTTTATTATTATATTTTATTATATTACA 33679

Qy        224 ATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTG 283
              | | | | |||     ||  ||||    |   |   | | ||  | ||||||     ||
Db      33678 TTTTATTATATTTCATATTATTTTTTATTTATTTTTTGTTTTATTTTATTCTATATTTT 33619

Qy        284 TTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTT 343
              ||  ||    || |  | || || | | |    ||   ||       || | ||| |||||
Db      33618 TTTTTTCATTTTATTTTATTATTTAGTTTAGTTTTATTATATTTTGTTATTTTTTGTTTT 33559

Qy        344 GTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATT 403
              || ||| | ||    |  |||  |    ||  || | |  |  |  | | ||     ||
Db      33558 GTTTTGTTATTGTATTTTATTTTTATTTTTTTATTCATTCGTTTATTCTTTTTTCGGCTT 33499

Qy        404 TGTT 407
              |||
Db      33498 AGTT 33495


RESULT 14
AC141780
LOCUS       AC141780               127172 bp    DNA     linear   HTG 19-MAR-2003
DEFINITION  Apis mellifera clone CH224-59E13, *** SEQUENCING IN PROGRESS ***,
            51 unordered pieces.
ACCESSION   AC141780
VERSION     AC141780.1  GI:29123964
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Apis mellifera (honeybee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 127172)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127172)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: AMCV
                Center clone name: CH224-59E13
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 109239 bases at least Q40
                Consensus quality: 114958 bases at least Q30
                Consensus quality: 118059 bases at least Q20
                Estimated insert size: 102169; sum-of-contigs estimation
                Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 51 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1115: contig of 1115 bp in length
*     1116     1215: gap of unknown length
*     1216     2262: contig of 1047 bp in length
*     2263     2362: gap of unknown length
*     2363     3978: contig of 1616 bp in length
*     3979     4078: gap of unknown length
*     4079     5145: contig of 1067 bp in length
*     5146     5245: gap of unknown length
*     5246     6692: contig of 1447 bp in length
*     6693     6792: gap of unknown length
*     6793     8155: contig of 1363 bp in length
*     8156     8255: gap of unknown length
*     8256     9653: contig of 1398 bp in length
*     9654     9753: gap of unknown length
*     9754    11173: contig of 1420 bp in length
*    11174    11273: gap of unknown length
*    11274    12683: contig of 1410 bp in length
*    12684    12783: gap of unknown length
*    12784    14617: contig of 1834 bp in length
*    14618    14717: gap of unknown length
*    14718    16311: contig of 1594 bp in length
*    16312    16411: gap of unknown length
*    16412    17991: contig of 1580 bp in length
*    17992    18091: gap of unknown length
*    18092    19407: contig of 1316 bp in length
*    19408    19507: gap of unknown length
*    19508    20717: contig of 1210 bp in length
*    20718    20817: gap of unknown length
*    20818    21864: contig of 1047 bp in length
*    21865    21964: gap of unknown length
*    21965    24010: contig of 2046 bp in length
*    24011    24110: gap of unknown length
*    24111    25260: contig of 1150 bp in length
*    25261    25360: gap of unknown length
*    25361    26988: contig of 1628 bp in length
*    26989    27088: gap of unknown length
*    27089    28652: contig of 1564 bp in length
*    28653    28752: gap of unknown length
*    28753    29941: contig of 1189 bp in length
*    29942    30041: gap of unknown length
*    30042    31428: contig of 1387 bp in length
*    31429    31528: gap of unknown length
*    31529    33482: contig of 1954 bp in length
*    33483    33582: gap of unknown length
*    33583    35392: contig of 1810 bp in length
*    35393    35492: gap of unknown length
*    35493    37241: contig of 1749 bp in length
*    37242    37341: gap of unknown length
*    37342    38797: contig of 1456 bp in length
*    38798    38897: gap of unknown length
*    38898    40712: contig of 1815 bp in length
*    40713    40812: gap of unknown length
*    40813    43143: contig of 2331 bp in length
*    43144    43243: gap of unknown length
*    43244    45600: contig of 2357 bp in length
*    45601    45700: gap of unknown length
*    45701    48432: contig of 2732 bp in length
*    48433    48532: gap of unknown length
*    48533    50872: contig of 2340 bp in length
*    50873    50972: gap of unknown length
*    50973    52456: contig of 1484 bp in length
*    52457    52556: gap of unknown length
*    52557    54380: contig of 1824 bp in length
*    54381    54480: gap of unknown length
*    54481    56322: contig of 1842 bp in length
*    56323    56422: gap of unknown length
*    56423    59508: contig of 3086 bp in length
*    59509    59608: gap of unknown length
*    59609    61786: contig of 2178 bp in length
*    61787    61886: gap of unknown length
*    61887    64028: contig of 2142 bp in length
*    64029    64128: gap of unknown length
*    64129    67129: contig of 3001 bp in length
*    67130    67229: gap of unknown length
*    67230    70067: contig of 2838 bp in length
*    70068    70167: gap of unknown length
*    70168    73427: contig of 3260 bp in length
*    73428    73527: gap of unknown length
*    73528    77057: contig of 3530 bp in length
*    77058    77157: gap of unknown length
*    77158    80255: contig of 3098 bp in length
*    80256    80355: gap of unknown length
*    80356    83534: contig of 3179 bp in length
*    83535    83634: gap of unknown length
*    83635    86553: contig of 2919 bp in length
*    86554    86653: gap of unknown length
*    86654    90241: contig of 3588 bp in length
*    90242    90341: gap of unknown length
*    90342    93387: contig of 3046 bp in length
*    93388    93487: gap of unknown length
*    93488    98046: contig of 4559 bp in length
*    98047    98146: gap of unknown length
*    98147   103358: contig of 5212 bp in length
*   103359   103458: gap of unknown length
*   103459   106861: contig of 3403 bp in length
*   106862   106961: gap of unknown length
*   106962   111298: contig of 4337 bp in length
*   111299   111398: gap of unknown length
*   111399   116662: contig of 5264 bp in length
*   116663   116762: gap of unknown length
*   116763   127172: contig of 10410 bp in length.
FEATURES             Location/Qualifiers
     source          1. .127172
                     /organism="Apis mellifera"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7460"
                     /clone="CH224-59E13"
ORIGIN

  Query Match           11.8%;  Score 49.2;  DB 2;  Length 127172;
  Best Local Similarity   52.0%;  Pred. No. 0.42;
  Matches  132;  Conservative    0;  Mismatches  120;  Indels    2;  Gaps    1;

Qy         94 TTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCC 153
              |||||     ||| ||| | ||  |     ||||| ||  ||  |      ||||||  |
Db      12152 TTAGGTATTTTTATTTTTTTTTTATTTNTATTATTTTTGTTGGTATAGTAATGATTGTTT 12211

Qy        154 GTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTC 213
              | |   | ||| | | |       | | |  |   || | |  ||  |     |||  |
Db      12212 ATTTGTTGTATTTTGTTAATTATATATTTGAGTTTTTATAATATATTATTATTATATGTAT 12271

Qy        214 TTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTAT 273
              ||  | |    || | |   ||| | |  | |||  | || | | | |||||  || |
Db      12272 TTATTTTTTTATGTTTTGGATTGTTGTGATTTTAGATATTTATTATTAGATATTTTTATA 12331

Qy        274 TCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTT 333
              |       || ||  | ||  ||   | |||| |||||| ||    | |  |||  | |||||
Db      12332 TTTTTTTTTATTAAATATTGTAAGGTTTTTTTTGATTGTGGATTAGGTATT--GTATTTT 12389

Qy        334 CTTTGGTTTTGTGT 347
              |||  ||||||  ||
Db      12390 ATTTNGTTTTGAGT 12403


RESULT 15
AX344573
LOCUS       AX344573               349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 24 from Patent WO0200932.
ACCESSION   AX344573
VERSION     AX344573.1  GI:18492459
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
           artificial sequences.
REFERENCE  1
  AUTHORS  Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE    Diagnosis of known genetic parameters within the mhc
  JOURNAL  Patent: WO 0200932-A 24 03-JAN-2002;
           Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo
                     sapiens)~Original length of seq 1: 3.673778 <223>~split as
                     follows.~seq 01 0.000.001 TO 0.349.980~seq 02 0.300.001
                     649.980~seq 03 600.001 949.980~seq 04 900.001
                     1.249.980~seq 05 1.200.001 1.549.980~seq 06 1.500.001
                     1.849.980~seq 07 1.800.001 2.149.980~seq 08 2.100.001
                     2.449.980~seq 09 2.400.001 2.749.980~seq 10 2.700.001
                     3.049.980~seq 11 3.000.001 3.349.980~seq 12 3.300.001
                     3.649.980~seq 13 3.600.001 3.673.778 <223>~Original length
                     of seq 2: 3.673778 <223>~split as follows.~seq 14
                     0.000.001 TO 0.349.980~seq 15 0.300.001 649.980~seq 16
                     600.001 949.980~seq 17 900.001 1.249.980~seq 18 1.200.001
                     1.549.980~seq 19 1.500.001 1.849.980~seq 20 1.800.001
                     2.149.980~seq 21 2.100.001 2.449.980~seq 22 2.400.001
                     2.749.980~seq 23 2.700.001 3.049.980~seq 24 3.000.001
                     3.349.980~seq 25 3.300.001 3.649.980~seq 26 3.600.001
                     3.673.778"
ORIGIN

  Query Match             11.8%;  Score 49.2;  DB 6;  Length 349980;
  Best Local Similarity   45.2%;  Pred. No. 0.37;
  Matches  180;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy         10 ATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGG 69
              | | || | |  | |||  ||||  |    | ||| |||  |  |   | ||  ||
Db     317812 AATTTTTGTTGTTTTTTCGTTTTTTTTTTTTAAATTATTTTTGGTAGTATTAATATATGAA 317871

Qy         70 AATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCT 129
              |||    |   ||| |  ||  |  | |     ||||  ||| | || ||    ||| | |
Db     317872 AATTATTTTTTAGTTTTTAGTATTTAATTTTTGATTTTTTTTTTTTTTTTTTAGATTTTTT 317931

Qy        130 TGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGC 189
                ||| ||       | |   |   || ||   ||||   ||  |  |      | |
Db     317932 ATTTTTTTTTTTTTTAAAAAAATAAATAAAATTTTTTTTATTTTTAAGTTTTAGTTTTTTT 317991

Qy        190 TAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCA 249
              ||  ||| ||       ||  |||  |  | | | || | |    | |    ||  ||
Db     317992 TATTTTTTATTTATTTATTTTTTGGTATTTTGGTATTTTTAGTAGTAAGTATTATTAATT 318051

Qy        250 CATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGAT 309
               ||  | | || ||| || ||| |    || | || | |    | |   |||| ||  |
Db     318052 TATTAATTATTAAATTTTTTTTAATTTAAGAGTATTAAGTTTAATGTAATTTTTTTTTTT 318111

Qy        310 TATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGC 369
              |    |     | || |  |||   |   | || | | |  | ||     || |||  ||
Db     318112 TGAAGGTAATTTAATATGTTTTTGGATAAATATTTTTTGTTTTTTTATTTATATTTTTGT 318171

Qy        370 AATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
               ||  ||    |    || |   | ||  |  ||| ||
Db     318172 TTTTTGATTTTTTAAAATATTTTTTTTTATTTTTTTT 318209


Search completed: June 20, 2004, 03:45:45
Job time : 2934.58 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 23:55:13 ; Search time 309.833 Seconds
(without alignments)
5731.297 Million cell updates/sec

Title: US-09-000-062-6
Perfect score: 418
Sequence: 1 tgaggtacgattcttcgatc.....taattttgttgaacagatccc 418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	418	100.0	418	2	AAT85996	Aat85996 Arabidope
2	54.4	13.0	5998	6	ABL33634	Abl33634 Human imm
3	50.6	12.1	17527	6	ABL33432	Abl33432 Human imm
4	50.6	12.1	17527	6	AAS63332	Aas63332 Chemicall
5	50.2	12.0	6306	4	AAS45515	Aas45515 Chemicall
6	50.2	12.0	6306	6	ABK28457	Abk28457 DNA trans
7	49	11.7	5987	6	ABL33562	Abl33562 Human imm
8	49	11.7	5987	6	ABQ67101	Abq67101 Human ang
9	49	11.7	5987	9	ADB54307	Adb54307 Pretreate
10	49	11.7	5987	9	ADB54179	Adb54179 Pretreate
11	48.8	11.7	15479	6	ABK39964	Abk39964 Human che
12	48.4	11.6	17848	4	AAS45323	Aas45323 Chemicall
13	48.4	11.6	17848	6	ABK39976	Abk39976 Human che
14	48.4	11.6	17848	6	ABK28164	Abk28164 DNA trans
15	48	11.5	2501	7	ABZ10074	Abz10074 Haematopo
16	48	11.5	2501	7	ABZ10220	Abz10220 Haematopo
17	48	11.5	6059	6	ABL33480	Abl33480 Human imm
18	48	11.5	8245	4	AAS46448	Aas46448 Tumour su
19	48	11.5	16724	6	ABL33090	Abl33090 Human imm
20	48	11.5	16724	6	ABL34536	Abl34536 Human met
21	48	11.5	16724	6	ABL70259	Abl70259 Chemicall
22	47.6	11.4	5649	9	ADB54254	Adb54254 Pretreate
23	47.6	11.4	5649	9	ADE84184	Ade84184 Human lym
24	47.4	11.3	21537	6	ABL33999	Abl33999 Human imm
25	47.2	11.3	5371	6	ABL34295	Abl34295 Human imm
26	46.8	11.2	714	6	ABQ18144	Abq18144 Oligonucl
c 27	46.8	11.2	714	6	ABQ18145	Abq18145 Oligonucl
28	46.8	11.2	8805	6	ABK40016	Abk40016 Human che
29	46.6	11.1	6012	6	ABL34564	Abl34564 Human met
30	46.6	11.1	6012	6	ABL70289	Abl70289 Chemicall
31	46.6	11.1	6668	6	ABL33697	Abl33697 Human imm
32	46.2	11.1	6048	6	ABL32509	Abl32509 Human imm
33	46.2	11.1	18817	6	ABL34495	Abl34495 Human met
34	46.2	11.1	18817	6	ABL70162	Abl70162 Chemicall
35	46.2	11.1	26811	2	AAX20253	Aax20253 Borrelia
36	46	11.0	5649	4	AAS46384	Aas46384 Tumour su
37	46	11.0	5649	6	ABK40008	Abk40008 Human che
38	46	11.0	5649	6	ABL32849	Abl32849 Human imm
39	46	11.0	5649	9	ADB54126	Adb54126 Pretreate
40	46	11.0	5649	9	ADE84108	Ade84108 Human lym
41	46	11.0	6161	6	ABL32623	Abl32623 Human imm
42	46	11.0	7892	6	ABK40056	Abk40056 Human che
43	45.8	11.0	6223	6	AAS61176	Aas61176 Human gen
44	45.6	10.9	374	7	ABX47736	Abx47736 Bovine ES
c 45	45.6	10.9	593	6	ABQ20161	Abq20161 Oligonucl

ALIGNMENTS

RESULT 1
AAT85996
ID AAT85996 standard; DNA; 418 BP.
XX
AC AAT85996;
XX
DT 17-NOV-1997 (first entry)
XX
DE Arabidopsis thaliana histone H3.3-like DNA fragment (intron 1).
XX
KW Plant expression regulation sequence; intron 1; histone;
KW herbicide tolerance; 5-enolpyruvylshikimate-3-phosphate synthase; EPSPS;
KW glyphosate; ds.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT intron 1. .418
FT /*tag= a
FT /number= 1
XX
PN WO9704114-A2.
XX
PD 06-FEB-1997.
XX
PF 17-JUL-1996; 96WO-FR001109.
XX
PR 19-JUL-1995; 95FR-00008980.
XX
PA (RHON) RHONE POULENC AGROCHIMIE.
XX
PI Derose R, Chaubet N, Gigot C;
XX
DR WPI; 1997-132652/12.
XX
PT New regulatory sequence for chimeric gene expression in rapidly growing
PT parts of a plant - includes at least one intron from a plant histone gene
PT and is useful for imparting resistance to herbicides.
XX
PS Claim 4; Page 26; 31pp; French.
XX
CC The known cosmid clone c22 of Arabidopsis thaliana contains two histone
CC H3.3-like genes. Digestion of clone c22 with restriction enzyme DdeI,
CC Klenow fragment and MsuI generated a fragment of 418 bp having the
CC present sequence. This fragment, designated intron 1, was ligated to
CC synthetic linkers for cloning into plant expression vectors. In addition

```
CC   to the intron 1 sequence, the vectors contained a plant promoter and a
CC   herbicide tolerance gene (e.g. a mutated version of the 5-
CC   enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene). The intron
CC   enhances expression of the herbicide tolerance gene in rapidly growing
CC   parts of plants. The intron can also be used to enhance expression of
CC   genes that impart resistance to pathogens or that encode nutritional or
CC   therapeutic proteins
XX
SQ   Sequence 418 BP; 100 A; 63 C; 85 G; 170 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 418;  DB 2;  Length 418;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-92;
  Matches  418;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418


RESULT 2
ABL33634
ID   ABL33634 standard; DNA; 5998 BP.
XX
AC   ABL33634;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1607.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1607; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 5998 BP; 1217 A; 92 C; 1806 G; 2883 T; 0 U; 0 Other;

  Query Match           13.0%;  Score 54.4;  DB 6;  Length 5998;
  Best Local Similarity  49.7%;  Pred. No. 0.0013;
  Matches  165;  Conservative    0;  Mismatches  166;  Indels    1;  Gaps    1;

Qy         78 GATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTT 137
              ||   |||||  ||   |||   |  |||||||| | | || |   |||  |  ||| ||
Db       4382 GAGTAGTGGTTTGAGTATAGTTAATTGTAGTTTCGAATTTTTGGGTTTAAATGATTTTTT 4441

Qy        138 CGCCGAATGATCTTCCGTA-TAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTT 196
              |    |  | ||  || |  |||| |||||||| || |   | || ||| ||  |
Db       4442 TGTTTTAGTTTTTTGATTAGTTGGGATTATAGGTTATAGTTATGTTGTTTAGTTAATTAA 4501

Qy        197 CATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGT 256
              | | | |  ||| ||| | || |      ||| || |     ||| |  |    ||
Db       4502 AAAAAAAAATTGTTTTTTTTTTTTTATAGAGATAGAAGTTTTTTTATGTTGTTTAGGTTGG 4561

Qy        257 AAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGA 316
              |  ||| ||  | || |||||||| ||   |    |||| | ||| |   ||||   |
Db       4562 TTTTGAATTTTTGGTTTTAGGCGATTTTTTTATTTTTTTTTTAGTTTTTGTGTTATTATA 4621

Qy        317 TTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAA 376
              ||   |       ||||  || ||| | | ||  | ||   ||   ||  |  |  |
Db       4622 TTTTTAGGGTAATTTTTTATTTGTTATTTATTATTTTTTGTATTTTTTTTTTTAGGTTTT 4681

Qy        377 TACGTATGGATCTAAATCTTGTTAATTTGTTG 408
              || |      || |     || |||  || | |
Db       4682 TATGGGATTATTTTTTGTTTTTTATGTTTTAG 4713


RESULT 3
ABL33432
ID   ABL33432 standard; DNA; 17527 BP.
XX
AC   ABL33432;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1405.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
```

PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1405; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 17527 BP; 4501 A; 296 C; 4094 G; 8636 T; 0 U; 0 Other;

Query Match 12.1%; Score 50.6; DB 6; Length 17527;
Best Local Similarity 52.1%; Pred. No. 0.013;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 192 GATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACA 251
Db 16590 GTTGTTATAAGAATATTTTGTTTTTTATTTGAGTCGTTATTATTTTAGAAAGTTTTAAGA 16649

Qy 252 TGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTA 311
Db 16650 GGTGGTATTTGTTGGTTGTTTTTTGTTGGTTTTTTTTTTTTTATTTTTTTTTTTTTAGTTT 16709

Qy 312 TGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAA 371
Db 16710 TAATAGTGATTATTGGGAAGTTTTTTTTTTGTTTTTGTTTGTTTTTTTTGTATTTTTTTAT 16769

Qy 372 TTGAATACGTATGGATCTAAATCTTGTTAATTTGTTG 408
Db 16770 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAG 16806

RESULT 4
AAS63332
ID AAS63332 standard; DNA; 17527 BP.
XX
AC AAS63332;
XX
DT 29-JAN-2002 (first entry)
XX
DE Chemically pretreated metabolism associated gene #27.
XX
KW Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
KW single nucleotide polymorphism detection; SNP; stool; urine; lung;
KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast; DUSP2;
KW EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.
XX
OS Homo sapiens.
XX
PN WO200176451-A2.
XX

PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP004016.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010834/01.
XX
PT New nucleic acid, useful for diagnosis and therapy of metabolic disease,
PT solid tumor and cancers, comprises segment of chemically modified genomic
PT sequences of genes associated with metabolism.
XX
PS Claim 1; Page 79-84; 143pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases of a segment of the chemically pretreated DNA of genes
CC associated with metabolism such as DUSP2 (NM_004418), EPHX2 (NM_001979),
CC QDPR (NM_000320), SGSH (NM_000199), SHMT2 (NM_005412), SLC7A2
CC (NM_003046), SLC7A4 (NM_004173) and TYMS (NM_001071) (all undefined). (I)
CC are useful for diagnosis and therapy of metabolic disease, solid tumours
CC and cancers; as primer oligonucleotides for the amplification of DNA
CC sequences, for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms (SNPs) in a chemically treated DNA of genes
CC associated with metabolism. An array of (I) is useful for ascertaining
CC genetic and/or epigenetic parameters for the diagnosis and/or therapy of
CC existing diseases or the predisposition to specific diseases by analysing
CC cytosine methylations. The method involves chemically treating genomic
CC DNA sample by a solution of bisulphite, hydrogen sulphite or disulphite
CC such that cytosine bases which are unmethylated at the 5th-position are
CC converted to uracil or another base which is dissimilar to cytosine in
CC terms of hybridisation behaviour and amplifying fragments of the
CC chemically pretreated genomic DNA. The genomic DNA is from cells or
CC cellular components which contain DNA, sources of DNA comprising, for
CC e.g. cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal
CC fluid, tissue embedded in paraffin such as tissue from eye, intestine,
CC kidney, brain, heart, prostate, lung, breast or liver, histologic object
CC slides and their combinations. Genetic parameters are mutations, in
CC particular insertions, deletions, point mutations, inversions and
CC polymorphisms of genes associated with metabolism and sequences further
CC required for their regulation. Epigenetic parameters are in particular
CC cytosine methylations and further chemical modifications of DNA bases of
CC genes associated with metabolism. Further epigenetic parameters include
CC for e.g. the acetylation of histones which correlates with DNA
CC methylation. AAS63306-AAS63373 represent chemically pretreated metabolism
CC associated genes, and related primers of the invention
XX
SQ Sequence 17527 BP; 4501 A; 296 C; 4094 G; 8636 T; 0 U; 0 Other;

Query Match 12.1%; Score 50.6; DB 6; Length 17527;
Best Local Similarity 52.1%; Pred. No. 0.013;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 192 GATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACA 251
Db 16590 GTTGTTATAAGAATATTTTGTTTTTTATTTGAGTCGTTATTATTTTAGAAAGTTTTAAGA 16649

Qy 252 TGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTA 311
Db 16650 GGTGGTATTTGTTGGTTGTTTTTTGTTGGTTTTTTTTTTTTTATTTTTTTTTTTTTAGTTT 16709

Qy 312 TGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAA 371
Db 16710 TAATAGTGATTATTGGGAAGTTTTTTTTTTGTTTTTGTTTGTTTTTTTTGTATTTTTTTAT 16769

Qy 372 TTGAATACGTATGGATCTAAATCTTGTTAATTTGTTG 408

```
Db      16770 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAG 16806


RESULT 5
AAS45515
ID   AAS45515 standard; DNA; 6306 BP.
XX
AC   AAS45515;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated complementary DNA associated with cell cycle #110.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002945.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analyzing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle.
XX
PS   Claim 1; SEQ ID NO 220; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers
XX
SQ   Sequence 6306 BP; 1455 A; 190 C; 1580 G; 3081 T; 0 U; 0 Other;

  Query Match             12.0%;  Score 50.2;  DB 4;  Length 6306;
  Best Local Similarity   50.2%;  Pred. No. 0.014;
  Matches  124;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy        161 GATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCT 220
              || ||| |||| |||  | || |||| | |||||||  | |  |  | |||| |  ||
Db       5490 GAGTTTCGGTTCGAGGGGTATAGTATGGTTAGATTTAGTGATTAATTTGTTTTTAGGTAG 5549

Qy        221 AGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGA 280
              |   | || | | |  |      ||||    | | ||  |   |  || || ||
Db       5550 ATGTTTTCGGTATTGATTTTAGATTTTTTGTTTTTTAGTTTATTGTTTTTTTTTTAAGTT 5609

Qy        281 TTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGT 340
              |  || | |    |||||  |    || ||    || ||    ||  |  ||||  |||  |
Db       5610 TGATTTAATTTATTTTTGGTGGGTTTTATATATCGTTTTTTTTTTTGTTTTTTTTTTTTT 5669

Qy        341 TTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTA 400
              ||| | || || ||   |   ||  ||   ||   |  | |  | |   | || ||
Db       5670 TTTTTTTTTATTTTTTTTTTTTTTTTTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5729

Qy        401 ATTTGTT 407
              ||| ||
Db       5730 TTTTTTT 5736


RESULT 6
ABK28457
ID   ABK28457 standard; DNA; 6306 BP.
XX
AC   ABK28457;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA transcription associated genomic DNA #166.
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW   neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW   myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW   angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW   polyglutamine disorder; solid tumour.
XX
OS   Unidentified.
XX
PN   WO200192565-A2.
XX
PD   06-DEC-2001.
XX
PF   06-APR-2001; 2001WO-EP003973.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-090046/12.
XX
PT   New nucleic acids or oligomers, useful for diagnosing or treating
PT   diseases associated with DNA transcription, e.g. immunological disorders,
PT   Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 331; 32pp; English.
XX
CC   The invention relates to a nucleic acid, which comprises a segment of the
CC   chemically pretreated DNA of genes associated with DNA transcription from
CC   one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC   or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC   to the chemically pretreated DNA of genes associated with DNA
CC   transcription. The set of oligomer probes are useful for detecting the
CC   cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC   in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC   diagnosing or treating diseases associated with DNA transcription
CC   (particularly with the methylation status), e.g. adenosine deaminase
```

CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 6306 BP; 1455 A; 190 C; 1580 G; 3081 T; 0 U; 0 Other;

Query Match 12.0%; Score 50.2; DB 6; Length 6306;
Best Local Similarity 50.2%; Pred. No. 0.014;
Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy       161 GATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCT 220
            || ||| |||| ||| | || |||| | ||||||| | | | | |||| | ||
Db      5490 GAGTTTCGGTTCGAGGGGTATAGTATGGTTAGATTTAGTGATTAATTTGTTTTTAGGTAG 5549

Qy       221 AGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGA 280
            |   | || | | |  |      |||| | | ||  |  | || || ||
Db      5550 ATGTTTTCGGTATTGATTTTAGATTTTTTGTTTTTTAGTTTATTGTTTTTTTTTTAAGTT 5609

Qy       281 TTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGT 340
            | || | |   |||| | |  || ||  || ||  ||  | |||| ||| |
Db      5610 TGATTTAATTTATTTTTGGTGGGTTTTATATATCGTTTTTTTTTTTGTTTTTTTTTTTT 5669

Qy       341 TTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTA 400
            ||| | || || ||   |  ||  ||  ||  | | | | |  | || ||
Db      5670 TTTTTTTTTATTTTTTTTTTTTTTTTTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5729

Qy       401 ATTTGTT 407
             ||| ||
Db      5730 TTTTTTT 5736
```

RESULT 7
ABL33562
ID ABL33562 standard; DNA; 5987 BP.
XX
AC ABL33562;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1535.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1535; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5987 BP; 1754 A; 96 C; 1096 G; 3041 T; 0 U; 0 Other;

Query Match 11.7%; Score 49; DB 6; Length 5987;
Best Local Similarity 47.5%; Pred. No. 0.026;
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```
Qy       108 TTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTA 167
            ||| | ||  ||   |  | | ||| ||     |||  | ||  | | ||  |||||
Db      2853 TTTTTTTTTATGTTTAGATTTGTATTTTTTTTTAAAATTTTGTTTATTTTTAAAGTTTTA 2912

Qy       168 GGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCT 227
            | ||      ||||  | |  | ||||  | |  | || |  | |||   | | | |
Db      2913 GATTTTTTTTGAATTTTTTTTTTTGATTTTTTAAGTATATGGAATTTTTTATTTTATTTT 2972

Qy       228 CTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGA 287
            | | ||   |||||  ||   ||      ||||||  |   ||| |       ||| |
Db      2973 GTTATATAAGTGATAGAATTTTTATTATGGGATAGATGGAGTTTAATTTTTTTGAGTTTA 3032

Qy       288 TTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGT 347
            |   |||  || |  ||  || |    ||    ||     || | ||||  |||  |
Db      3033 AAATAATTTAAATATAATTATTTTTTATGTTTTGTTTTTTTTTTTATTTTTGTATTTAAAT 3092

Qy       348 TGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
            |  | ||  ||||      |       |   || |||    |||  || | |||  ||
Db      3093 TTTTTTTTAGATAATAGAATAATTAATGTTTGATAAGGAAGATAATGATGATGATTATTT 3152

Qy       408 GAACA 412
             || |
Db      3153 TAAAA 3157
```

RESULT 8
ABQ67101
ID ABQ67101 standard; DNA; 5987 BP.
XX
AC ABQ67101;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 131.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX

PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 131; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5987 BP; 1754 A; 96 C; 1096 G; 3041 T; 0 U; 0 Other;

Query Match 11.7%; Score 49; DB 6; Length 5987;
Best Local Similarity 47.5%; Pred. No. 0.026;
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 108 TTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTA 167
Db 2853 TTTTTTTTTATGTTTAGATTTGTATTTTTTTTTAAAATTTTGTTTATTTTTAAAGTTTTA 2912

Qy 168 GGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCT 227
Db 2913 GATTTTTTTTGAATTTTTTTTTTTGATTTTTAAGTATATGGAATTTTTTATTTTATTTT 2972

Qy 228 CTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGA 287
Db 2973 GTTATATAAGTGATAGAATTTTTATTATGGGATAGATGGAGTTTAATTTTTTGAGTTTA 3032

Qy 288 TTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGT 347
Db 3033 AAATAATTTAAATATAATTATTTTTTATGTTTTGTTTTTTTTTTTATTTTTGTATTTAAAT 3092

Qy 348 TGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
Db 3093 TTTTTTTTAGATAATAGAATAATTAATGTTTGATAAGGAAGATAATGATGATGATTATTT 3152

Qy 408 GAACA 412
Db 3153 TAAAA 3157

RESULT 9
ADB54307
ID ADB54307 standard; DNA; 5987 BP.
XX
AC ADB54307;
XX
DT 04-DEC-2003 (first entry)
XX
DE Pretreated genomic DNA region 231.
XX
KW colon cell proliferative disorder; non methylated CpG dinucleotide;
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX

PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX
PT Detecting and differentiating between colon cell proliferative disorders
PT associated with a gene or its regulatory regions comprises contacting a
PT target nucleic acid in a biological sample obtained from the subject with
PT a reagent.
XX
PS Claim 32; SEQ ID NO 363; 74pp; English.
XX
CC The invention relates to a novel method for detecting and differentiating
CC between colon cell proliferative disorders associated with at least one
CC gene or its regulatory regions. The method comprises contacting a target
CC nucleic acid in a biological sample obtained from the subject with at
CC least one reagent or a series of reagents, where the reagent or series of
CC reagents, distinguishes between methylated and non methylated CpG
CC dinucleotides within the target nucleic acid. The molecules of the
CC invention demonstrate cytostatic activity whilst the method may useful
CC for detecting and differentiating between colon cell proliferative
CC disorders, including cancers such as colon adenoma and colon carcinoma.
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC determining cytosine methylation state or single nucleotide
CC polymorphisms. The current sequence is that of the pretreated genomic DNA
CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 5987 BP; 1754 A; 0 C; 1096 G; 3137 T; 0 U; 0 Other;

Query Match 11.7%; Score 49; DB 9; Length 5987;
Best Local Similarity 47.5%; Pred. No. 0.026;
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 108 TTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTA 167
Db 2853 TTTTTTTTTATGTTTAGATTTGTATTTTTTTTTAAAATTTTGTTTATTTTTAAAGTTTTA 2912

Qy 168 GGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCT 227
Db 2913 GATTTTTTTTGAATTTTTTTTTTTGATTTTTAAGTATATGGAATTTTTTATTTTATTTT 2972

Qy 228 CTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGA 287
Db 2973 GTTATATAAGTGATAGAATTTTTATTATGGGATAGATGGAGTTTAATTTTTTGAGTTTA 3032

Qy 288 TTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGT 347
Db 3033 AAATAATTTAAATATAATTATTTTTTATGTTTTGTTTTTTTTTTTATTTTTGTATTTAAAT 3092

Qy 348 TGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
Db 3093 TTTTTTTTAGATAATAGAATAATTAATGTTTGATAAGGAAGATAATGATGATGATTATTT 3152

Qy 408 GAACA 412
Db 3153 TAAAA 3157

RESULT 10
ADB54179
ID ADB54179 standard; DNA; 5987 BP.
XX

```
AC   ADB54179;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Pretreated genomic DNA region 103.
XX
KW   colon cell proliferative disorder; non methylated CpG dinucleotide;
KW   cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS   Unidentified.
XX
PN   WO2003072821-A2.
XX
PD   04-SEP-2003.
XX
PF   27-FEB-2003; 2003WO-EP002035.
XX
PR   27-FEB-2002; 2002EP-00004551.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Adorjan P,  Burger M,  Maier S,  Nimmrich I,  Becker E,  Lesche R;
PI   Rujan T,  Schmitt A;
XX
DR   WPI; 2003-731620/69.
XX
PT   Detecting and differentiating between colon cell proliferative disorders
PT   associated with a gene or its regulatory regions comprises contacting a
PT   target nucleic acid in a biological sample obtained from the subject with
PT   a reagent.
XX
PS   Claim 32; SEQ ID NO 235; 74pp; English.
XX
CC   The invention relates to a novel method for detecting and differentiating
CC   between colon cell proliferative disorders associated with at least one
CC   gene or its regulatory regions. The method comprises contacting a target
CC   nucleic acid in a biological sample obtained from the subject with at
CC   least one reagent or a series of reagents, where the reagent or series of
CC   reagents, distinguishes between methylated and non methylated CpG
CC   dinucleotides within the target nucleic acid. The molecules of the
CC   invention demonstrate cytostatic activity whilst the method may useful
CC   for detecting and differentiating between colon cell proliferative
CC   disorders, including cancers such as colon adenoma and colon carcinoma.
CC   The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC   determining cytosine methylation state or single nucleotide
CC   polymorphisms. The current sequence is that of the pretreated genomic DNA
CC   region of the invention. This sequence is not shown within the
CC   specification but is taken from Wipoweb.
XX
SQ   Sequence 5987 BP; 1754 A; 96 C; 1096 G; 3041 T; 0 U; 0 Other;

  Query Match             11.7%;  Score 49;  DB 9;  Length 5987;
  Best Local Similarity   47.5%;  Pred. No. 0.026;
  Matches  145;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy        108 TTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTA 167
              ||| | ||  ||    |  | | ||| ||      |||  | ||   | | ||  |||||
Db       2853 TTTTTTTTTATGTTTAGATTTGTATTTTTTTTTAAAATTTTGTTTATTTTTAAAGTTTTA 2912

Qy        168 GGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCT 227
              | ||     |||||  | |   | |||||  | |  | || |  | |||    |  | | |
Db       2913 GATTTTTTTTGAATTTTTTTTTTTGATTTTTTAAGTATATGGAATTTTTTATTTTATTTT 2972

Qy        228 CTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGA 287
              | | ||    |||||  ||   ||      ||||||  |   ||| |          ||| |
Db       2973 GTTATATAAGTGATAGAATTTTTATTATGGGATAGATGGAGTTTAATTTTTTTGAGTTTA 3032

Qy        288 TTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGT 347
              |   |||  || |  ||  || |    ||     ||  || | |||| |||   |
Db       3033 AAATAATTTAAATATAATTATTTTTTATGTTTTGTTTTTTTTTTATTTTTGTATTTAAAT 3092

Qy        348 TGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
              |  | ||  ||||       |          |   || |||    |||  || | |||  ||

Db       3093 TTTTTTTTAGATAATAGAATAATTAATGTTTGATAAGGAAGATAATGATGATGATTATTT 3152

Qy        408 GAACA 412
                 | | |
Db       3153 TAAAA 3157


RESULT 11
ABK39964
ID   ABK39964 standard; DNA; 15479 BP.
XX
AC   ABK39964;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #23 strand 1.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 45; 24pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence at least 18
CC   bases in length of a segment of the chemically pretreated DNA of genes
CC   associated with pharmacogenomics according to one of the sequences of the
CC   genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC   (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC   (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC   NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC   their complementary sequences, or a sequence (S1) chosen from 87
CC   sequences and their complements. The chemical pretreatment is bisulphite
CC   treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC   Also included are an oligomer (II) in particular an oligonucleotide or a
CC   peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC   base sequence having a length of 9 nucleotides which hybridises to or is
CC   identical to a chemically pretreated DNA of genes associated with
CC   pharmacogenomics and their complements, arranged in an array for
CC   analysing diseases associated with the methylation state (CpG) and/or
CC   detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC   oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC   and their complements is useful for diagnosis and therapy of solid
CC   tumours and cancer. The present sequence represents one the 87 DNA
CC   sequences or its complement. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 15479 BP; 4303 A; 122 C; 2850 G; 8204 T; 0 U; 0 Other;

  Query Match             11.7%;  Score 48.8;  DB 6;  Length 15479;
  Best Local Similarity   51.9%;  Pred. No. 0.034;
```

```
  Matches  110;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        204 AGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGA 263
              || |||||| ||||| |  ||     |     | |   | |||  |  ||| ||    |
Db      11257 AGGTAGTTTTTTTGTTTTTAAATATTTTTTGTATTTTTTTTTTATATTTGTATATGGTGT 11316

Qy        264 TTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAAT 323
              || |  |  || |   || ||||||     ||  || || || |||||     ||  |
Db      11317 TTTTATTATTTAGAATGTTTTTGATTTTTTATTATATGTTTTTTATTATTATTGTGGTTT 11376

Qy        324 TAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTAT 383
              |||  | ||  ||  | ||   ||| | ||  |     |||        ||   || | |
Db      11377 GGGGGTTGTTGGTTTTTGTTTGTTTGTTTTTTGGTGGTATTTTGATTTTTATTTAAGATT 11436

Qy        384 GGATCTAAATCTTGTTAATTTGTTGAACAGAT 415
              | | || || || || |||| || || ||||
Db      11437 TGGTTTAGATATTATTTATTTTTTAAAAAGAT 11468


RESULT 12
AAS45323
ID   AAS45323 standard; DNA; 17848 BP.
XX
AC   AAS45323;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated complementary DNA associated with cell cycle #14.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002945.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analyzing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle.
XX
PS   Claim 1; SEQ ID NO 28; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers
XX
SQ   Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;

  Query Match           11.6%;  Score 48.4;  DB 4;  Length 17848;
  Best Local Similarity   46.9%;  Pred. No. 0.044;
  Matches  151;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy         91 AAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCT 150
              ||||       |  ||| | | | || |     | |   || ||| ||      ||
Db       1226 AAATATTTTAAAGTTAATATATTTTTTAAATGAATTAATTATTTTTTTTATTAAATTGAG 1285

Qy        151 TCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGT 210
              |  | |  |  |||||| ||  ||   |   ||||        || ||    |||  |
Db       1286 TTTTTTTTTGTATTTTAGATTTTAGGGTATGAATATATTATTTATTTATTGATAGAAGTT 1345

Qy        211 TTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCT 270
              || ||| | |  ||| | |  | || | |  ||||  || |||   || ||| ||
Db       1346 TTTTTTTTGTTAAATTTTAGTGGTTATTAAATTTTTTTGTATTTGTTTTTATATTTTTTA 1405

Qy        271 TATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGAT 330
              ||||  |    || ||  |||   || |  || | || |  ||  | ||   ||||   |
Db       1406 TATTTGTTTTTTTTTTTTTATTTTATTTTTATGTTTAAGATTAGTTTTTTATTATTTTT 1465

Qy        331 TTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTA 390
              | | |||      | || | | ||     | |||  ||   | | | |  | |||  |
Db       1466 TGTTTTTTTGATGTTGGTTTTGTTTTTTTTAATTTATGTTTTATATTGTGGTTAGATTGA 1525

Qy        391 AATCTTGTTAATTTGTTGAACA 412
              |      |||| |  |  | |
Db       1526 ATATGAAATAATATTGTTTATA 1547


RESULT 13
ABK39976
ID   ABK39976 standard; DNA; 17848 BP.
XX
AC   ABK39976;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #29 strand 2.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 58; 24pp; English.
XX
```

CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;

Query Match 11.6%; Score 48.4; DB 6; Length 17848;
Best Local Similarity 46.9%; Pred. No. 0.044;
Matches 151; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 91 AAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCT 150
Db 1226 AAATATTTTAAAGTTAATATATTTTTTAAATGAATTAATTATTTTTTTTATTAAATTGAG 1285

Qy 151 TCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGT 210
Db 1286 TTTTTTTTTGTATTTTAGATTTTAGGGTATGAATATATTATTTATTTATTGATAGAAGTT 1345

Qy 211 TTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCT 270
Db 1346 TTTTTTTTGTTAAATTTTAGTGGTTATTAAATTTTTTTGTATTTGTTTTTATATTTTTTA 1405

Qy 271 TATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGAT 330
Db 1406 TATTTGTTTTTTTTTTTTTATTTTTATTTTTATGTTTAAGATTAGTTTTTTATTATTTTT 1465

Qy 331 TTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTA 390
Db 1466 TGTTTTTTTTGATGTTGGTTTTGTTTTTTTTAATTTATGTTTTATATTGTGGTTAGATTGA 1525

Qy 391 AATCTTGTTAATTTGTTGAACA 412
Db 1526 ATATGAAATAATATTGTTTATA 1547

RESULT 14
ABK28164
ID ABK28164 standard; DNA; 17848 BP.
XX
AC ABK28164;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated complementary genomic DNA #19.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP003973.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
PS Claim 1; SEQ ID NO 38; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;

Query Match 11.6%; Score 48.4; DB 6; Length 17848;
Best Local Similarity 46.9%; Pred. No. 0.044;
Matches 151; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

Qy 91 AAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCT 150
Db 1226 AAATATTTTAAAGTTAATATATTTTTTAAATGAATTAATTATTTTTTTTATTAAATTGAG 1285

Qy 151 TCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGT 210
Db 1286 TTTTTTTTTGTATTTTAGATTTTAGGGTATGAATATATTATTTATTTATTGATAGAAGTT 1345

Qy 211 TTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCT 270
Db 1346 TTTTTTTTGTTAAATTTTAGTGGTTATTAAATTTTTTTGTATTTGTTTTTATATTTTTTA 1405

Qy 271 TATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGAT 330

```
Db       1406 TATTTGTTTTTTTTTTTTTATTTTTATTTTTATGTTTAAGATTAGTTTTTTATTATTTTT 1465

Qy        331 TTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTA 390

Db       1466 TGTTTTTTTGATGTTGGTTTTGTTTTTTTTAATTTATGTTTTATATTGTGGTTAGATTGA 1525

Qy        391 AATCTTGTTAATTTGTTGAACA 412

Db       1526 ATATGAAATAATATTGTTTATA 1547


RESULT 15
ABZ10074
ID   ABZ10074 standard; DNA; 2501 BP.
XX
AC   ABZ10074;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #214.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 214; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX

SQ   Sequence 2501 BP; 532 A; 124 C; 707 G; 1138 T; 0 U; 0 Other;

  Query Match             11.5%;  Score 48;  DB 7;  Length 2501;
  Best Local Similarity   47.1%;  Pred. No. 0.04;
  Matches  180;  Conservative    0;  Mismatches  200;  Indels    2;  Gaps    1;

Qy         25 TTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGT 84

Db       1661 TTTATATTTTTTGAGATTTAGTTTCGGTTTAAATAGTTTGATTTTAAATTGTTGAAGGGA 1720

Qy         85 GGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAA 144

Db       1721 AAAAGGAAAGTTGTTGTTATTATTTCGTTTTTTTGTGTTTTAGTTTTTTTAATTGTGAAA 1780

Qy        145 TGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCA 204

Db       1781 TGGAGTTCGTAGTGTTTTTTATTTGAGAGTTATGAAAGTTTTTATAAGTTTTTAGTTTAT 1840

Qy        205 GATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGAT 264

Db       1841 AATAGTTGTGTTAGTTATTATTATGATGATGATTACGATATGTTATTTTTTTGGGGAAAA 1900

Qy        265 TGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATT 324

Db       1901 TAATATTTTTTTGTTTAGTAGTAGTATTGGTTTTTTTTTTTT--TCGTTATTATTAAATT 1958

Qy        325 AGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATG 384

Db       1959 TTGGTTTTATTTTATTTTGGTTTAGTTTTTTTTATAGGTGTATTTGTTTGTAGGGGGTTA 2018

Qy        385 GATCTAAATCTTGTTAATTTGT 406

Db       2019 GATATTATTGTTTTTGTTTTTT 2040


Search completed: June 20, 2004, 01:58:56
Job time : 312.833 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:39:08 ; Search time 2071.67 Seconds
(without alignments)
6025.290 Million cell updates/sec

Title: US-09-000-062-6
Perfect score: 418
Sequence: 1 tgaggtacgattcttcgatc.....taatttgttgaacagatccc 418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	379.8	90.9	436	28	BH633573	BH633573 SALK_0427
c	2	353.2	84.5	366	28	BH846820	BH846820 SALK_0105
c	3	194.4	46.5	258	29	AL950572	AL950572 Arabidops
c	4	192.4	46.0	262	29	AL950573	AL950573 Arabidops
c	5	66	15.8	223	29	AL950952	AL950952 Arabidops
	6	56.6	13.5	1180	13	BX436369	BX436369 BX436369
c	7	54.6	13.1	1180	9	AL565108	AL565108 AL565108
c	8	54.4	13.0	750	29	CNS011ID	AL100303 Drosophil
	9	53.4	12.8	1201	9	AL524807	AL524807 AL524807
c	10	52.6	12.6	960	29	CNS005F3	AL059925 Drosophil
c	11	52.4	12.5	832	29	CNS0081L	AL051208 Drosophil
	12	51.8	12.4	712	13	BX416727	BX416727 BX416727
c	13	51.6	12.3	1101	29	CNS00Z6Z	AL097301 Drosophil
c	14	50.2	12.0	894	29	CNS018BG	AL109126 Drosophil
c	15	50.2	12.0	975	29	CNS014RT	AL104531 Drosophil
c	16	50	12.0	1020	29	CNS038G9	AL232578 Tetraodon
c	17	49.8	11.9	878	29	CNS0187R	AL108993 Drosophil
	18	49.8	11.9	1124	13	BX436282	BX436282 BX436282
c	19	49.4	11.8	953	13	BX331505	BX331505 BX331505
	20	49.2	11.8	524	29	CNS01U9O	AL167541 Tetraodon
	21	49.2	11.8	997	13	BX419578	BX419578 BX419578
c	22	48.8	11.7	783	29	CNS00AIS	AL055833 Drosophil
	23	48.8	11.7	1015	13	BX399782	BX399782 BX399782
	24	48.6	11.6	843	29	AG177964	AG177964 Pan trogl
c	25	48.4	11.6	695	28	BZ462901	BZ462901 BONRH10TF
	26	48.4	11.6	897	29	CNS07ABZ	AL436389 T7 end of
	27	48.4	11.6	1201	13	BX403954	BX403954 BX403954
	28	48.2	11.5	888	29	CNS005I9	AL059534 Drosophil
c	29	48.2	11.5	964	29	CNS006N9	AL065781 Drosophil
c	30	48.2	11.5	1101	29	CNS00LJT	AL068307 Drosophil
c	31	48.2	11.5	1201	13	BX446430	BX446430 BX446430
c	32	48	11.5	759	29	CNS06QXV	AL411257 T7 end of
	33	48	11.5	1043	29	CNS0145P	AL103735 Drosophil
	34	48	11.5	1157	29	CNS02UGV	AL214456 Tetraodon
	35	48	11.5	1201	9	AL554111	AL554111 AL554111
	36	47.8	11.4	1101	29	CNS008XS	AL052569 Drosophil
	37	47.8	11.4	1101	29	CNS012NV	AL101797 Drosophil
c	38	47.6	11.4	1068	29	CNS00J4X	AL075933 Drosophil
	39	47.6	11.4	1101	29	CNS00D7M	AL060328 Drosophil
c	40	47.6	11.4	1101	29	CNS00EVL	AL069706 Drosophil
c	41	47.4	11.3	996	29	CNS00DIX	AL070681 Drosophil
c	42	47.4	11.3	1165	13	BX338369	BX338369 BX338369
	43	47.2	11.3	1047	28	CC266341	CC266341 CH261-12K
c	44	47.2	11.3	1061	13	BX437039	BX437039 BX437039
c	45	47.2	11.3	1101	29	CNS00B37	AL056366 Drosophil

ALIGNMENTS

RESULT 1
BH633573/c
LOCUS BH633573 436 bp DNA linear GSS 01-FEB-2002
DEFINITION SALK_042781 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_042781, genomic survey sequence.
ACCESSION BH633573
VERSION BH633573.1 GI:18468936
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 436)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At4g40040.
Class: TDNA tagged.

FEATURES Location/Qualifiers
source 1. .436
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_042781"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 90.9%; Score 379.8; DB 28; Length 436;
Best Local Similarity 99.5%; Pred. No. 1.6e-67;
Matches 381; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         33 CCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGTGGTACGAA 92
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        435 CCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGTGGTACGAA 376

Qy         93 ATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTC 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        375 ATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTC 316

Qy        153 CGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTT 212
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        315 CGGATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTT 256

Qy        213 CTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTA 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTA 196

Qy        273 TTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTT 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 TTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTT 136

Qy        333 TCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAA 392
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        135 TCTTTGGTTTTGTGTTGATCTTACGATACATTCCGGCAATTGAATACGTATGGATCTAAA 76

Qy        393 TCTTGTTAATTTGTTGAACAGAT 415
              |||||||||||||||||||||||
Db         75 TCTTGTTAATTTGTTGAACAGAT 53
```

RESULT 2
BH846820/c
LOCUS BH846820 366 bp DNA linear GSS 13-JUN-2002
DEFINITION SALK_010583.54.50.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_010583.54.50.x, genomic survey sequence.
ACCESSION BH846820
VERSION BH846820.1 GI:21417691
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 366)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At4g40040.
Class: TDNA tagged.

FEATURES Location/Qualifiers
source 1. .366
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_010583.54.50.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN

Query Match 84.5%; Score 353.2; DB 28; Length 366;
Best Local Similarity 97.8%; Pred. No. 4.1e-62;
Matches 358; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy         30 TTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGTGGTAC 89
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 TTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGTGGTAC 307

Qy         90 GAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATC 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 GAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATC 247

Qy        150 TTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAG 209
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 TTCCGGATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAG 187

Qy        210 TTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTC 269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 TTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTC 127

Qy        270 TTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGA 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGA 67

Qy        330 TTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCT 389
              ||||||| ||| | |||||||||||||||||||||| ||||| ||||||||| |||| |
Db         66 TTTTCTTCGGTATCGTGTTGATCTTACGATACATTCCGGCAATCGAATACGTAGGGATGT 7

Qy        390 AAATCT 395
              ||||||
Db          6 AAATCT 1
```

RESULT 3
AL950572/c
LOCUS AL950572 258 bp DNA linear GSS 24-OCT-2002
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-329E10-016035, genomic survey sequence.
ACCESSION AL950572
VERSION AL950572.1 GI:24407194
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

```
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 258)
  AUTHORS   Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At4g40040. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .258
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-329E10-016035"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             46.5%;  Score 194.4;  DB 29;  Length 258;
  Best Local Similarity   99.5%;  Pred. No. 8.4e-30;
  Matches  195;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        220 TAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 TAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCG 177

Qy        280 ATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGG 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 ATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGG 117

Qy        340 TTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTT 399
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        116 TTTTGTGTTGATCTTACGATACATTCCGGCAATTGAATACGTATGGATCTAAATCTTGTT 57

Qy        400 AATTTGTTGAACAGAT 415
              ||||||||||||||||
Db         56 AATTTGTTGAACAGAT 41


RESULT 4
AL950573/c
LOCUS       AL950573                 262 bp    DNA     linear   GSS 24-OCT-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-329E10-016044,
            genomic survey sequence.
ACCESSION   AL950573
VERSION     AL950573.1  GI:24407195
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 262)
  AUTHORS   Strizhov,N., Li,Y., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At4g40040. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .262
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-329E10-016044"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             46.0%;  Score 192.4;  DB 29;  Length 262;
  Best Local Similarity   99.5%;  Pred. No. 2.1e-29;
  Matches  193;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        220 TAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 TAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCG 203

Qy        280 ATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGG 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 ATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGG 143

Qy        340 TTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTT 399
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        142 TTTTGTGTTGATCTTACGATACATTCCGGCAATTGAATACGTATGGATCTAAATCTTGTT 83

Qy        400 AATTTGTTGAACAG 413
              ||||||||||||||
Db         82 AATTTGTTGAACAG 69


RESULT 5
AL950952/c
LOCUS       AL950952                 223 bp    DNA     linear   GSS 24-OCT-2002
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence GK-332E10-016059,
            genomic survey sequence.
```

```
ACCESSION   AL950952
VERSION     AL950952.1  GI:24407574
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.
            and Weisshaar,B.
  TITLE     A pipeline for automated high-throughput generation of FSTs
            (flanking sequence tags) from Arabidopsis thaliana T-DNA
            transformed lines
  JOURNAL   Unpublished
REFERENCE   2
  AUTHORS   Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  TITLE     A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
            for flanking sequence tag based reverse genetics
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 223)
  AUTHORS   Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
            Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
COMMENT     This sequence is recovered from the left border of the T-DNA. It
            indicates an insertion close to or within gene At4g40040. The
            sequences are generated at the MPI for Plant Breeding Research in
            the context of the GABI-Kat project. GABI-Kat is part of the German
            Plant Genomics program designated 'GABI'. Information on line
            availability can be found at:
            http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES             Location/Qualifiers
     source          1. .223
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="GK-332E10-016059"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
                     /note="PCR was performed on DNA from Arabidopsis thaliana
                     plants (T1) which were transformed with the T-DNA from
                     vector pAC161. The lines contain one or more T-DNA
                     insertions. The DNA fragment(s) resulting from the PCR
                     were directly sequenced to determine the genomic sequence
                     flanking the insertion. Sequences displaying significant
                     similarity to the A. thalaina nuclear genome sequence were
                     processed for submission. T-DNA derived sequences were
                     removed"
ORIGIN

  Query Match             15.8%;  Score 66;  DB 29;  Length 223;
  Best Local Similarity   87.8%;  Pred. No. 0.0011;
  Matches   72;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy        220 TAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCG 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 TAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCT 164

Qy        280 ATTGTTGATTAGGGTTTTGATT 301
              || ||  |  ||||||  ||
Db        163 TTTCTTTTTGGGGGTTTCCCTT 142


RESULT 6
BX436369
LOCUS       BX436369               1180 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX436369 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YM12
            5-PRIME, mRNA sequence.
ACCESSION   BX436369
VERSION     BX436369.1  GI:30787529
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1180)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 516.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0CAP001BG06QP1&cluster=516.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP001BG06QP1.
FEATURES             Location/Qualifiers
     source          1. .1180
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YM12"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             13.5%;  Score 56.6;  DB 13;  Length 1180;
  Best Local Similarity   27.3%;  Pred. No. 0.067;
  Matches  111;  Conservative  101;  Mismatches  195;  Indels    0;  Gaps    0;

Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              || :::  :    :  : |  : ||| | ||  |      |  || |: :||: : : ::
Db        728 TGKKKKNKKKRGTKVNKKTTTWTTTTTTTRTTTTTTTTTTTTRTTRTKDKGTKKKTKGRM 787

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
                :  :::    : : ::|  ::|::::     ::  : : |:|:  ||| | || ||
Db        788 TKKKKGKKRKKRKKKKTKDTTKKTKKKDDTTKKKKTTKKKKGVTKGTTTTTTTTTTTTTT 847

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
                || | ||:|| |: :  :  :: : ::     :  ::  |:| |: |:    :
Db        848 TTTTTTTTTTKTTTGTKGKKKKKGKKKKKKKKGNGTKKKDKKRTKTKGKGTVTKNTKNTK 907

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              | :: ::  |   ::|  |     : |  :|: :||:  :    | | |::  || |
Db        908 TKKKTKDNTTKTTKKTDNTTKKTTKTTKTKTDTKTTKRTKTTTTTKTTTDMTTTTTTKTT 967

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              |  ||||   :|::||   : : |:::: :: :  :: : || ||: :     ||:|: :
Db        968 TTTTTTTTTTMTKKGTKTTKKKNTKKKKTKKKKNTKKTKKTTTTTKTKNTTTTTTKTKTK 1027

Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||  ||: || |    ||    ||  :  |||| |||:::::::  ||: :|||    |
Db       1028 TTCTTTKTTTKTTTTTTTTKBTTTTTSTTTTTTTTTKKKKKKKSTTTKTKCTTTTTTTT 1087

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
              ||  |   ::   |  | | : | |   : ||:||  ||: ||
Db       1088 TTTTTTTTTTKBKTTCTTKTTBTTKTTTTTTBKTTTSTTTKTTBYTT 1134


RESULT 7
AL565108/c
LOCUS       AL565108               1180 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL565108 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone
            CS0DN004YK09 3-PRIME, mRNA sequence.
ACCESSION   AL565108
```

```
VERSION     AL565108.2  GI:31289080
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1180)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 16, 2001 this sequence version replaced gi:12916155.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 7664.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DN004AF05NP1&cluster=7664.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DN004AF05NP1.
FEATURES             Location/Qualifiers
     source          1. .1180
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DN004YK09"
                     /tissue_type="ADULT BRAIN"
                     /dev_stage="adult"
                     /clone_lib="Homo sapiens ADULT BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN
```

```
  Query Match             13.1%;  Score 54.6;  DB 9;  Length 1180;
  Best Local Similarity   36.6%;  Pred. No. 0.17;
  Matches  106;  Conservative   52;  Mismatches  132;  Indels    0;  Gaps    0;

Qy        104 TTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGAT 163
              ||  ||| | || ||      || | || ||| || :  |: | ::  |    : |:| ||
Db        690 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTWDAWGDTTTWHWVTATARWAAWAWAT 631

Qy        164 TTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGA 223
              |:||  |  : : |   | :: |:  :|   :  |  |  |:||| ||      : :: :
Db        630 TWTATATGTKTKYTTTKTTRWTTWAAWAWTAWAAAAAAAWWADATAATTADAAAAWWWRWW 571

Qy        224 ATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTG 283
              :: :    |||| | :|:|  :|:  ::|: : : |||: ::    ::||      ||
Db        570 WDTWTATTAATTRTTKAWAARWTWWTMWTRADAWTATADTWWTAAAWWATARTTTTATAT 511

Qy        284 TTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTT 343
              ||  ||     ||:| || ||:|:: |  |     |||   :|:    |||| |||  ||||
Db        510 TTTTTTTTTTTTYTTATGTTMTWRTTMTTTTTTTTGTTTKTRTATTTTTTTTTTTTTTTT 451

Qy        344 GTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAAT 393
              | ||: | ||    || :|:  :    :|  :|: :|||  || |: |:
Db        450 TTTTTSTTTTTTATTTATWTDTTKAKTKDTWGRTRTRTATANATATDKAK 401
```

```
RESULT 8
CNS011ID/c
LOCUS       CNS011ID                750 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN06D21 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL100303
VERSION     AL100303.1  GI:5611914
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN06D21"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN
```

```
  Query Match             13.0%;  Score 54.4;  DB 29;  Length 750;
  Best Local Similarity   32.1%;  Pred. No. 0.2;
  Matches  127;  Conservative   74;  Mismatches  195;  Indels    0;  Gaps    0;

Qy         12 TCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAA 71
              || |   |  | |||  |||| |       |  |::|:  :|  |  | |:: :  :|:
Db        530 TCATTTTTTTTTTTTTTTTTTCCTTTTTTTATKKTWTTKTTTTTWTTAWWTYTTWGKTT 471

Qy         72 TCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTG 131
              :  |   |:::: |       | ||   |  || |:|| | :: ||     :| || :
Db        470 WGTTTTTTTWKKWATTTKTTTTTAAGWTTATKTTTGKTTTTTYKTTTTTTTTKTTTCAKT 411

Qy        132 TTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTA 191
              :|| :|    || |  : ||  :: |    ||::   |:  :  | | | :| ||| |
Db        410 KTTGYTTTTKGATTTTKTTTTTKWTTYTTTTTTWKTTTTWTTWTCTAAGTTKTYTAGTTT 351

Qy        192 GATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACA 251
              |:|  |:     ||  | : ||| | |   |  ::  : || :  | |||:
Db        350 TCTYTATTYCTTTAATKWTAKTTTTTTWTCTTKTAGYWTTWTTTGWTTTTKTTTTKWCTG 291

Qy        252 TGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTA 311
              | : :  ::   :| |: || |: :    ||::   :  : |||| :: ||   ||
Db        290 TAKTKTKWKTTTYTCTWTTTTTKWKTAWTTTKKYTTAWTAKCTTTTKTTTTTKKWTKTTT 231

Qy        312 TGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAA 371
              |   |||   :||   : :||: ::|  |:| :| ||    :| |  |  :||: |:
Db        230 TTTTATTAKTWTTCCTKTWTTKWKWTTTTWTAKTTTTCGAKKTTCTTTTTWTTYTTKKGT 171

Qy        372 TTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
              |::  |  || : | | |:  :  |   |:::||
Db        170 TWKYKTTTGTTYTGTTTTWKTKKTCTTWTTTWKKTT 135
```

```
RESULT 9
AL524807
LOCUS       AL524807                1201 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  AL524807 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC008YK01 3-PRIME, mRNA sequence.
ACCESSION   AL524807
VERSION     AL524807.2  GI:31043067
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12788300.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3928.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC008AF01NP1&cluster=3928.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DC008AF01NP1.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC008YK01"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match 12.8%; Score 53.4; DB 9; Length 1201;
Best Local Similarity 42.9%; Pred. No. 0.3;
Matches 129; Conservative 23; Mismatches 149; Indels 0; Gaps 0;

```
Qy        104 TTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGAT 163
              ||  ||| | || ||   | || | || ||| ||     :||  | :|   |||  :  :
Db        203 TTTTTTTTTCTTTTTTTTCTTTTTTTTTTTTTTTTTTTWATTTTWTTTTTATTTWTTW 262

Qy        164 TTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGA 223
              |||   :|     | :::  : ||  ::  |||| |       |  ||| ||| | ||
Db        263 TTTTTTWTTTTTTTTWWWATWTTATTWWWTTTTCTTTTTTTTTTTTTTTTTTTTTTTTT 322

Qy        224 ATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTG 283
              | | | : ||: | : | ::|||    | | |   | : ||:|| || ||      ||
Db        323 TTTTTTTWTATKGTTKTTTKKTTTTTTKTTTTTTTTTTKKTTKTTTTTTTTTTTTTTTT 382

Qy        284 TTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTT 343
              ||  ||    ||||  |||| ||  |: |  : |||   ||     |||| |||  ||||
Db        383 TTTTTTTTTTTTTTTTTTTTTTTTTTTKTTTKTTGTTTTTTTTTTTTTTTTTTTTTTT 442

Qy        344 GTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATT 403
              : ||  | ||    |  ||  :||  ||  |   |  | |  | || ||  ||
Db        443 WYTTTTTTTTTTTTTTTTTTTTTTKGCCTTTTTTTTTTTTAGTTTTTTKTTTTTTTTTTT 502

Qy        404 T 404
              :
Db        503 W 503
```

RESULT 10
CNS005P3/c
LOCUS CNS005F3 960 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR11P07 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL059925
VERSION AL059925.1 GI:4943047
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 960)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .960
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR11P07"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 12.6%; Score 52.6; DB 29; Length 960;
Best Local Similarity 40.2%; Pred. No. 0.45;
Matches 158; Conservative 30; Mismatches 205; Indels 0; Gaps 0;

```
Qy         15 TCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGAATCG 74
              | ||   : |:|| | ||  : :   | ||| :|  :   | :||  |  : ||:    |
Db        572 TWGARDTKTTKTGTTGTTTTKTKTGTTTTTTGWTTTKGTGTTKTGTGAAARKTGRKTGTG 513

Qy         75 CTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTT 134
                ::|  |||: :  ||:||  |  :|  |  ||| | || ||:|    :|| :|| :|
Db        512 TGWRWTGTGTGRADAAAAWTTKTGAWDGKGTGTTTTGTTTTKTTKGTGTGWATGKTGRDT 453

Qy        135 CTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGAT 194
              | :   ::  : ||   |:|   || |||: :|| ||::::   | :  || |   |
Db        452 KGTKKTGTGRDTGWTTTTGTTDTGKTGAATTTDTKTTKGADWWTTTTGTKKAAGRTTTTT 393

Qy        195 TTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGT 254
              ||  |      |  ||| ||| | |    | | |    || |   |  ||||    | |
Db        392 TTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 333

Qy        255 GTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGC 314
              |   |   || || || ||      ||  |   |    ||||  |||| ||  || |
Db        332 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTNTTTTTTTTTTTTTTTTTTTTTTTT 273

Qy        315 GATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTG 374
              ||    ||     |||| |||  |||| | ||  | ||    |  ||  |    ||
Db        272 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 213

Qy        375 AATACGTATGGATCTAAATCTTGTTAATTTGTT 407
              |   | |   | |     || ||  ||| ||
Db        212 TTTTTTTTTTTTTTTTTNTTTTTTTTTTTTT 180
```

RESULT 11
CNS0081L/c
LOCUS CNS0081L 832 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR16N08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

```
ACCESSION   AL051208
VERSION     AL051208.1  GI:4933161
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .832
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR16N08"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             12.5%;  Score 52.4;  DB 29;  Length 832;
  Best Local Similarity   39.4%;  Pred. No. 0.51;
  Matches   86;  Conservative   38;  Mismatches   94;  Indels    0;  Gaps    0;

Qy        190 TAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCA 249
              ||  |:|    :  | ||: |:| |:| | |::  : | ::: :|| |   |  |||| :
Db        278 TACKTWTATATWWTATATWTTWTKTWTTTTTWKKCKTTWWKWKWTTATWTTTTTTTTTGW 219

Qy        250 CATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGAT 309
                |   | | : :| || || || |    || || |:|    ||||  |||| ||  |
Db        218 TTATTTCTATTTKTWTTTTTTTTTTTGTTTCTTCTTTAYTTTTTTTTTTTTTTTTWTTTTT 159

Qy        310 TATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGC 369
              | |   | |  : :|:   :|||| ||| |||| :::|  | |:   |  |:::  |
Db        158 TTTTTTAATATWTWTWTWTTTTWTTTTTTTTTTWKKWTTTTTTKTTTACCCWKKKWTTW 99

Qy        370 AATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
              :| ::   || |  |||||| | ||:::  |:::::
Db         98 TTKTAWWCYKGTTTCTATCTAAKTWTTKKWTKTYKKKW 61


RESULT 12
BX416727
LOCUS       BX416727                712 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA011YI14 5-PRIME, mRNA sequence.
ACCESSION   BX416727
VERSION     BX416727.1  GI:30765629
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DA011BE07QP1.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA011YI14"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             12.4%;  Score 51.8;  DB 13;  Length 712;
  Best Local Similarity   10.8%;  Pred. No. 0.69;
  Matches   33;  Conservative  152;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        104 TTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGAT 163
              ||  ||| |:|| ||      || | ||:|:| ||       |  | ||  :| |   : |
Db        377 TTTTTTTTWTTTTTTTTTTTTTTTTKTKTKTTTTTTTTTTTTTTTTTKTTTTKTKTT 436

Qy        164 TTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGA 223
              ::: ::|: : : :   : :| : : |   |::        : | :::: :::|: : :
Db        437 KKKTKKTKKKKKTKTTTKTKTKKKKKTKTKTKKKTGKKGKKKKTKKKKKKKKKGKGKKKK 496

Qy        224 ATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTG 283
               : : ::   :: : : : ::::    :::::   : : ::::: :: :: :: : :::
Db        497 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKDKKKKKKKKKKKKKKKKKKKKKKKKK 556

Qy        284 TTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTT 343
              :::|:: :::::::: :::: ::: :: :: : :::   :: ::: :::: :::::::::
Db        557 KKKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 616

Qy        344 GTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATT 403
              :::::: : ::  : :   ::  ::   ::: :  :: ::: : :   : :::::  ::
Db        617 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 676

Qy        404 TGTTG 408
              :::::
Db        677 KKKKK 681


RESULT 13
CNS00Z6Z/c
LOCUS       CNS00Z6Z                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN01A10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097301
VERSION     AL097301.1  GI:5608912
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
```

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN01A10"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

ORIGIN

Query Match 12.3%; Score 51.6; DB 29; Length 1101;
Best Local Similarity 33.0%; Pred. No. 0.7;
Matches 136; Conservative 81; Mismatches 189; Indels 6; Gaps 1;

```
Qy          2 GAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGAA 61
              :::  ::  ||::|  : : | ||| ::|::  |: ::     | |  :| : : :: :
Db        954 KVRMVMMMTTTVWTWNDMHDTGTTTTMHTKDKKTRNMRGNGGGTGTGTVGGKKKKKKTHK 895

Qy         62 ACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGC 121
                |:  |:   :| |  ||:  |  |:  :: : :   : :|:  |:: |:: :||
Db        894 NGTRTVGRSKGMGTTDAATWATTTWTRWWWWTWAWWAWTWAMTKTWTWKKTHWAMTTAHR 835

Qy        122 CATTATCTT------GTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAG 175
              ::|::| ||      ||:| | ::   ||          :| |||: ||: :| |:  |
Db        834 MMTWWTTTTTTKTGAGTWTTTRMVAAAAAAATAAAATAWWAAAAAVTTTWAMGWTWTWAA 775

Qy        176 ATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATT 235
              |  |||  |:||   : |   |: | :  : | :|: ||::|   :|| | |  :  |
Db        774 AAAAATATTWTAWTAGRKTAATAHDAAMTAWRAWWTWWTTKVTKMTAWATTTTTTTVTMT 715

Qy        236 CTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTT 295
              :|:  |:  |||:  :| ::|||| : |:|:|: :  :|   :  |:    | ||   ||
Db        714 YTMTTTDTGTTTTMTAMTMKKTAAAAMAAKTKTWTKMTKTKAAMWTTKAMGAAATAATTTT 655

Qy        296 TTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTA 355
              ||  | ||  | |||||   |:|  | :|     ||||  |||  |||| : ||  | |:
Db        654 TTTTTATTTATTATTATATAAWTATAYYTTTTTTTTTTTTTTTTTTTTTYTTTTCTTTYT 595

Qy        356 CGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
                |  ||  |  ||  |  | |  |  |||| ||  ||| ||
Db        594 TTTTTTTTTTTTTTTYTTTTTTTTTTTTTTTYTTTTTTCTTTTTTTTTTTT 543
```

RESULT 14
CNS018BG/c
LOCUS CNS018BG 894 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN13B16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL109126
VERSION AL109126.1 GI:5629430
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 894)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .894
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN13B16"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

ORIGIN

Query Match 12.0%; Score 50.2; DB 29; Length 894;
Best Local Similarity 38.1%; Pred. No. 1.4;
Matches 140; Conservative 39; Mismatches 188; Indels 0; Gaps 0;

```
Qy         43 TTTTTTCGGTGATCGTGAAACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAG 102
              ||||||| ||: | |  | | |  :  |  |  || ::  |  |  || |  |
Db        550 TTTTTTCTGTKTTTTTTTTAGTGCGTKTGTTCCAACKTATKYYTTKWTATGTTTGTKAAA 491

Qy        103 ATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGA 162
                |  |||  || | |  |: | ||  :|||| ||        |  | ||   |  :
Db        490 TCTTTTTTWGATACGTTTCYTTGATGCWGTTTATTTTTGTTTTTKTTTTTGTTKAWKGTT 431

Qy        163 TTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAG 222
              |:||   :|    ||   :  | : : ||  |||  |   |  :: :||| |:   :
Db        430 TWTTTTTKTTTKAATCYKWWTTYKTKTTATTTTTTTTTKTCTWWKTKTTTTTYCCAGYTT 371

Qy        223 AATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATT 282
                | ||   || |   |  |||   | | |   |  || || || ||       ||
Db        370 TTGTTTTGTTTTTTTKTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 311

Qy        283 GTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTT 342
              ||  ||  || ||| ||| ||  || | |   ||    ||      |||| |||::|||
Db        310 TTTTTTTTTTGTGTTGTGTTTTTTTTTTTTTTTTTTTGTTTTTTTTTTTTTTTTKKTTT 251

Qy        343 TGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAAT 402
              |:|:::| | ||        || ::   :|:  ::|  :|:| :|  |   | :: :|::|
Db        250 TBTKYKGTTTTTKTTTGTWKTTTYYTTTWTKATWWAKTKAYGTWTTTTKKTTWKTWTWWT 191

Qy        403 TTGTTGA 409
                |  ||::
Db        190 TATTTKW 184
```

RESULT 15
CNS014RT/c
LOCUS CNS014RT 975 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN12L12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL104531
VERSION AL104531.1 GI:5616545
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 975)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

```
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN12L12"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             12.0%;  Score 50.2;  DB 29;  Length 975;
  Best Local Similarity   41.0%;  Pred. No. 1.4;
  Matches  103;  Conservative   28;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        103 ATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGA 162
              |||  | | | || || ||| :| : || ||| ||     :|: :: ::   | ::::
Db        253 ATTTTTATTTTTTTTTTGCCGWTTWTTTTTTTTTTTTTTTWWAWWWWWWWWWWTTWWWWYT 194

Qy        163 TTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAG 222
              :|||| || | | | : :   :   :|| : |  | | | ||| ||| ||||
Db        193 WTTTAAATTTCATTTACTTACWTWTTTWTWTWTTTTWAATTATAAAATTTTTTTTTCTAT 134

Qy        223 AATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATT 282
              | | |    |: :  | ::|||   : | |   |   ||  ||:| ||   |  ||
Db        133 TTTTTTTTTTTWTWWTTTTKKTTTCCTTWTTTTCCTTTTTTTTCTCKTTTTTTCCTCTT 74

Qy        283 GTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTT 342
              ||  ||    |::|  |||||||  ||:| |  ||    ||      | | :||  |||
Db         73 TTTYCTTTTTTTYYTTTTTTTCTTTTTTMTACTTTTTTTTTTTTTNTTMTNTTYTTNTTTT 14

Qy        343 TGTGTTGATCT 353
              | |  |  | |
Db         13 TTTTCTTCTNT 3


Search completed: June 20, 2004, 05:01:18
Job time : 2075.67 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:39:08 ; Search time 2448.33 Seconds
(without alignments)
6025.290 Million cell updates/sec

Title: US-09-000-062-7
Perfect score: 494
Sequence: 1 ctcaggcgaagaacaggtat.....cgcagatcccgggatctgcg 494

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	266.4	53.9	300	9	AV551146	AV551146 AV551146
	2	202.8	41.1	320	14	Z34742	Z34742 ATTS3502 St
	3	61.4	12.4	712	13	BX416727	BX416727 BX416727
c	4	54.4	11.0	1101	29	CNS012JN	AL101645 Drosophil
	5	54	10.9	576	29	CNS035N7	AL228940 Tetraodon
c	6	52.4	10.6	1131	14	CD050625	CD050625 AGENCOURT
c	7	51.8	10.5	683	29	CNS00Z8F	AL097353 Drosophil
	8	51.8	10.5	942	29	CNS018GS	AL109318 Drosophil
c	9	51.4	10.4	873	13	BU534806	BU534806 AGENCOURT
c	10	51.2	10.4	750	29	CNS011ID	AL100303 Drosophil
	11	51	10.3	1246	29	CG744146	CG744146 P036-3-E1
	12	50	10.1	741	29	CNS007YN	AL051001 Drosophil
	13	49.6	10.0	563	29	CNS007UZ	AL050869 Drosophil
	14	49.6	10.0	737	29	CNS008BU	AL051476 Drosophil
	15	49.6	10.0	1403	29	CG749401	CG749401 P043-3-E0
	16	49.4	10.0	1388	29	CG756681	CG756681 P051-4-D1
c	17	49.2	10.0	1022	13	BU529542	BU529542 AGENCOURT
c	18	49	9.9	1108	29	CNS007IO	AL067316 Drosophil
	19	49	9.9	1201	9	AL514421	AL514421 AL514421
	20	48.8	9.9	483	29	CNS02CV8	AL191645 Tetraodon
	21	48.8	9.9	848	12	BG809697	BG809697 mgct001xf
c	22	48.8	9.9	1761	28	CC188336	CC188336 CH261-70H
	23	48.6	9.8	1103	28	BZ551529	BZ551529 pacs1-60_
	24	48.6	9.8	1201	13	BX336467	BX336467 BX336467
	25	48.4	9.8	864	29	CNS06O5G	AL407642 T7 end of
c	26	48.4	9.8	1058	14	CD048666	CD048666 AGENCOURT
c	27	48.4	9.8	1300	12	BM468018	BM468018 AGENCOURT
	28	48.2	9.8	878	29	CNS00Z8X	AL097371 Drosophil
	29	48.2	9.8	1059	29	CNS00Z2B	AL097133 Drosophil
	30	48.2	9.8	1376	29	CG747831	CG747831 P041-3-B0
	31	47.8	9.7	922	29	CNS0073W	AL066784 Drosophil
	32	47.8	9.7	1180	13	BX436369	BX436369 BX436369
	33	47.8	9.7	1215	12	BM922560	BM922560 AGENCOURT
	34	47.6	9.6	446	12	BP516575	BP516575 BP516575
c	35	47.6	9.6	914	13	BU563375	BU563375 AGENCOURT
c	36	47.6	9.6	967	29	CNS077ZW	AL433362 T7 end of
c	37	47.4	9.6	614	29	CNS0152H	AL104915 Drosophil
	38	47	9.5	867	29	CNS0054A	AL057618 Drosophil
c	39	47	9.5	1053	13	BU508694	BU508694 AGENCOURT
	40	47	9.5	1289	29	CG748984	CG748984 P043-1-C0
	41	47	9.5	1345	29	CG746697	CG746697 P040-1-B0
c	42	46.8	9.5	767	29	CNS00AQX	AL055924 Drosophil
c	43	46.8	9.5	854	29	CNS012CM	AL101392 Drosophil
c	44	46.8	9.5	966	14	CD388802	CD388802 AGENCOURT
c	45	46.8	9.5	995	14	CD385043	CD385043 AGENCOURT

ALIGNMENTS

RESULT 1
AV551146
LOCUS AV551146 300 bp mRNA linear EST 06-SEP-2000
DEFINITION AV551146 Arabidopsis thaliana roots Columbia Arabidopsis thaliana cDNA clone RZ121h02R 5', mRNA sequence.
ACCESSION AV551146
VERSION AV551146.1 GI:8722559
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7 (3), 175-180 (2000)
MEDLINE 20363093
PUBMED 10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers

```
    source          1. .300
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /strain="Columbia"
                    /db_xref="taxon:3702"
                    /clone="RZ121h02R"
                    /tissue_type="roots"
                    /clone_lib="Arabidopsis thaliana roots Columbia"
                    /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                    XhoI"
ORIGIN

  Query Match             53.9%;  Score 266.4;  DB 9;  Length 300;
  Best Local Similarity   99.6%;  Pred. No. 4.5e-45;
  Matches  267;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 92

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 152

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        153 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 212

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 272

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGA 268
              ||||||||||||||||||||||||||||
Db        273 TCTTGCTTCGAAGTTCTGTATAACCAGA 300


RESULT 2
Z34742
LOCUS       Z34742                   320 bp    mRNA    linear   EST 11-JUL-1994
DEFINITION  ATTS3502 Strasbourg-A Arabidopsis thaliana cDNA clone FAI185 5',
            mRNA sequence.
ACCESSION   Z34742
VERSION     Z34742.1  GI:507089
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 320)
  AUTHORS   CNRS.
  TITLE     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Philipps G., Gigot C.
            Gigot Claude / L512
            Laboratoire de Biologie Moleculaire des Plantes - CNRS
            12 Rue du General Zimmer,67084 Strasbourg Cedex,France
            Email: ARABANK@MEDOC.U-STRASBG.FR.
FEATURES             Location/Qualifiers
     source          1. .320
                     /organism="Arabidopsis thaliana"
                     /mol_type="mRNA"
                     /strain="ecotype Columbia"
                     /db_xref="taxon:3702"
                     /clone="FAI185"
                     /clone_lib="Strasbourg-A"
                     /note="Vector: Lambda ZAPII; tissue=sliced leaves of
                     A.thaliana ecotype columbia; clone_library=Strasbourg-A;
                     Cloning vector: Lambda ZAPII; Physiological condition:
                     leaves strips incubated 2/3/4 days in liquid culture
                     medium."
ORIGIN

  Query Match             41.1%;  Score 202.8;  DB 14;  Length 320;
  Best Local Similarity   85.3%;  Pred. No. 5.5e-32;
  Matches  273;  Conservative    0;  Mismatches   42;  Indels    5;  Gaps    4;

Qy         17 GTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATG-TTT 75
              |||||||||||||||||||||||||| ||| ||  |   || ||||||||||||| |||
Db          1 GTATGATTTGTTTGTAATTAGATCAAGGGGTTTAGGTCTTTCATTACTTTTTAATGTTTT 60

Qy         76 TTTCTGTTACTGTCTCC--GCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCC 133
              |||||||||||||||||  |||||||||| ||||||||||||||||||||||||||||
Db         61 TTTCTGTTACTGTCTCCCGCGATCCTGATTTTACCGCAATAGAGTTTCGGGTTTTGTCCA 120

Qy        134 ATTCCAGTTTGAAAATAAACGT-CCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAA 192
              | | |||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db        121 TTCCAAGTTTGAAAATAAAGGTCCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAA 180

Qy        193 GATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAA 252
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        181 GATTGAGTCTAGTCGATTTATTGGATGA-CCATTCTTCATCGTTTTTTTCTTGCTTCGAA 239

Qy        253 GTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAAC 312
              |||||||||||||   || |   |   |  || ||||||||||||||||||| |
Db        240 GTTCTGTATAACCAAGATTTGCCTGTGTCCGATTGCAATACCTAGCCGTGTATCGAGACC 299

Qy        313 TAGGGTTTTCGAGTCAATTT 332
              ||||| |||||||  | |||
Db        300 TAGGGGTTTCGAGCAATTTT 319


RESULT 3
BX416727
LOCUS       BX416727                 712 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA011YI14 5-PRIME, mRNA sequence.
ACCESSION   BX416727
VERSION     BX416727.1  GI:30765629
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DA011BE07QP1.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA011YI14"
                     /tissue_type="NEUROBLASTOMA"
                     /clone_lib="Homo sapiens NEUROBLASTOMA"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             12.4%;  Score 61.4;  DB 13;  Length 712;
  Best Local Similarity   13.4%;  Pred. No. 0.0058;
  Matches   53;  Conservative  179;  Mismatches  165;  Indels    0;  Gaps    0;
```

```
Qy          9 AAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTT 68
Db        307 WAWWWAWWAWWWAWWWWTTWWATWWWTWWTTTTTTTTTTTTTTTTTTTTTTWTTTTTTT 366

Qy         69 AATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTT 128
Db        367 TTTTTTTTTTTTTTTTTTTTWTTTTTTTTTTTTTTTTTKTKTKTTTTTTTTTTTTTTTT 426

Qy        129 GTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTG 188
Db        427 KTTTTKTKTTKKKTKKTKKKKKTKTTTKTKTKKKKKTKTKTKKKTGKKGKKKKTKKKKK 486

Qy        189 TGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTT 248
Db        487 KKKGKGKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKDKKKKKKKKKKKKKKKKKK 546

Qy        249 CGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGA 308
Db        547 KKKKKKKKKKKKKAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 606

Qy        309 GAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTC 368
Db        607 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 666

Qy        369 ATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATT 405
Db        667 KKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 703
```

RESULT 4
CNS012JN/c
LOCUS CNS012JN 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN08O24 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL101645
VERSION AL101645.1 GI:5613256
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN08O24"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 11.0%; Score 54.4; DB 29; Length 1101;
Best Local Similarity 40.9%; Pred. No. 0.14;
Matches 168; Conservative 36; Mismatches 203; Indels 4; Gaps 1;

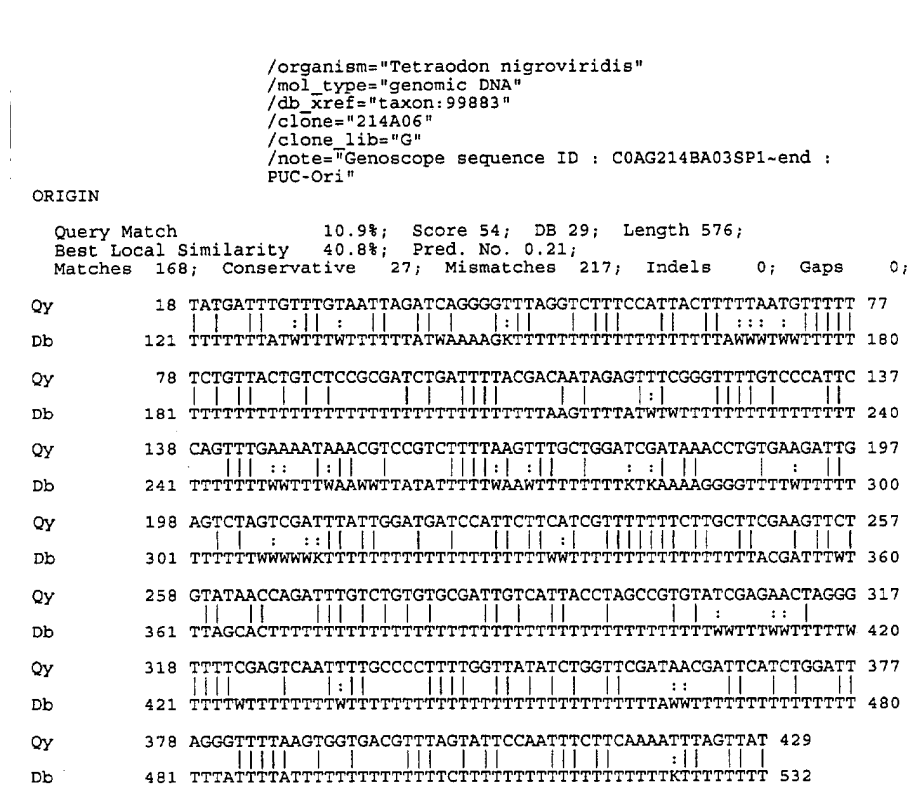
```
Qy         23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
Db        431 TYYKATTTTTTTWTTWTTTTTTTTTTTTTTTTTTTTTTYTTATTTTTTTATTTTTTTTTTT 372

Qy         83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
Db        371 TTTWTTTTTTTTTTTTTTTTWTTTCCWTKWTTTTTWGTTYTTWTTTTTTTTYCWTTKTWTTT 312

Qy        143 TGAAAATAAACGTCCGTCTTTTA----AGTTTGCTGGATCGATAAACCTGTGAAGATTGA 198
Db        311 TTTTTTAWTTTTAWTTTTTWTATACCTTTTTTCTTTTTTTTTTATACTYTTWWATWTTWT 252

Qy        199 GTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTG 258
Db        251 TTTCTTTTTTWTTTTTTATTTTTCTWTYATTTATTTTTTTWTTTTTCTYTCWAAGATCAG 192

Qy        259 TATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGT 318
Db        191 KWTAATTTWAGKTTTWAAAAAAAAAAGCGGAAAGTCCACTTAATTTATTTWTTWAWATTTT 132

Qy        319 TTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTA 378
Db        131 TTTTTATTWWWTTTTTTACYTTTTCCTTTTTWTTWAWTTTTATWWATTWATTTWTTANAA 72

Qy        379 GGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTAT 429
Db         71 GATTCCTTCCTCTTTTTTTTTTWTTTTTTTTTTTTTTTTTTTTTTTTTTT 21
```

RESULT 5
CNS035N7
LOCUS CNS035N7 576 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 214A06 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL228940
VERSION AL228940.1 GI:7887933
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 576)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .576

/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="214A06"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG214BA03SP1~end : PUC-Ori"

ORIGIN

Query Match 10.9%; Score 54; DB 29; Length 576;
Best Local Similarity 40.8%; Pred. No. 0.21;
Matches 168; Conservative 27; Mismatches 217; Indels 0; Gaps 0;



RESULT 6
CD050625/c
LOCUS CD050625 1131 bp mRNA linear EST 09-MAY-2003
DEFINITION AGENCOURT_13977661 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CD050625
VERSION CD050625.1 GI:30487745
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1131)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jamie Thompson, University of WI
cDNA Library Preparation: Gina Zastrow-Hayes
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDKM32 row: m column: 05
High quality sequence start: 66
High quality sequence stop: 257.
FEATURES Location/Qualifiers
source 1. .1131
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic trophoblasts, made from WA01 stem cells"
/lab_host="DH10B TonA"
/clone_lib="NIH_MGC_173"
/note="Vector: pDONR201; Site_1: attP2; Site_2: attP1; LIBR_PRIMING - oligo dT; METHOD - full-length enriched; LIBR_PROVIDER - Bradfield"

ORIGIN

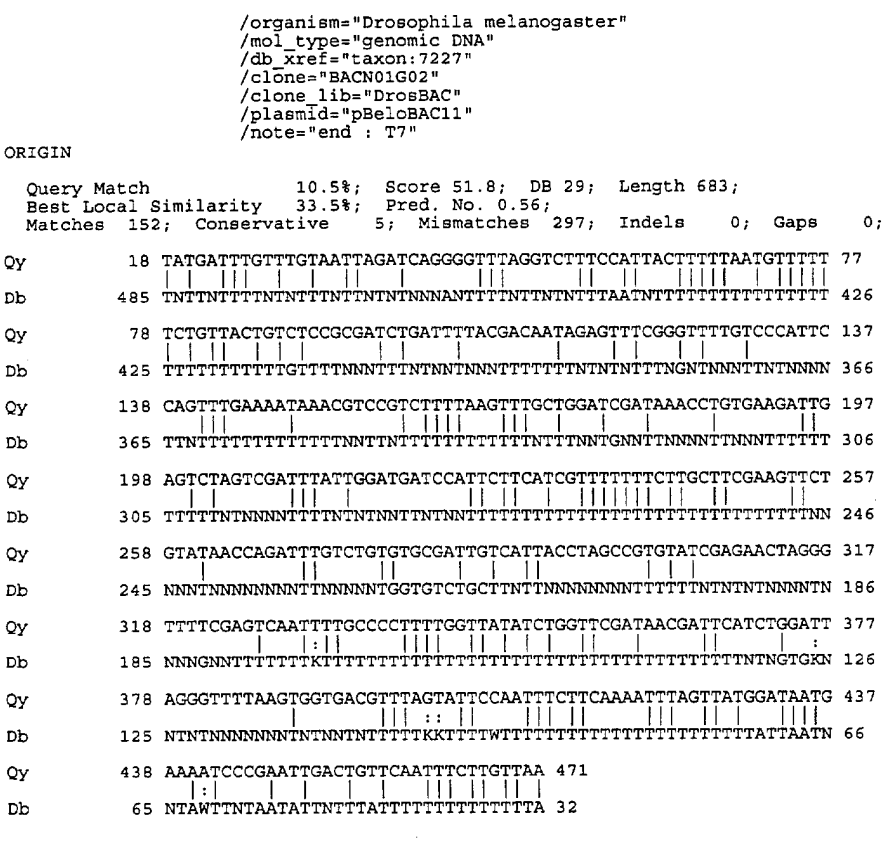
Query Match 10.6%; Score 52.4; DB 14; Length 1131;
Best Local Similarity 45.4%; Pred. No. 0.37;
Matches 188; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

RESULT 7
CNS00Z8F/c
LOCUS CNS00Z8F 683 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN01G02 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL097353
VERSION AL097353.1 GI:5608964
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 683)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .683

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN01G02"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

ORIGIN

Query Match 10.5%; Score 51.8; DB 29; Length 683;
Best Local Similarity 33.5%; Pred. No. 0.56;
Matches 152; Conservative 5; Mismatches 297; Indels 0; Gaps 0;



RESULT 8
CNS018GS
LOCUS CNS018GS 942 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN13P09 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL109318
VERSION AL109318.1 GI:5629622
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 942)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .942
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN13P09"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

ORIGIN

Query Match 10.5%; Score 51.8; DB 29; Length 942;
Best Local Similarity 36.3%; Pred. No. 0.51;
Matches 113; Conservative 48; Mismatches 150; Indels 0; Gaps 0;

RESULT 9
BU534806/c
LOCUS BU534806 873 bp mRNA linear EST 13-SEP-2002
DEFINITION AGENCOURT_10183661 NIH_MGC_143 Mus musculus cDNA clone IMAGE:6562130 5', mRNA sequence.
ACCESSION BU534806
VERSION BU534806.1 GI:22845247
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 873)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Michael Brownstein
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM2737 row: 1 column: 02
High quality sequence stop: 141.
FEATURES Location/Qualifiers
source 1. .873
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

/clone="IMAGE:6562130"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_143"
/note="Organ: Brain; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.2-0.5 kb size fraction (other fractions present in NIH_MGC_144). Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.4%; Score 51.4; DB 13; Length 873;
Best Local Similarity 32.3%; Pred. No. 0.63;
Matches 112; Conservative 0; Mismatches 235; Indels 0; Gaps 0;

```
Qy         23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
Db        582 TTTTTTTTTTTTTTTNTNNNNNNNNNNNNNNNTNNNNTNTTTTTTTTTTTTTTTTTTNT 523

Qy         83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
Db        522 NNTTTTTTTTTTTTTTTTNTTTNNNNNNNNNNTNTTTTTTNTTNTTTTTTNNTNNNNNTTT 463

Qy        143 TGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCT 202
Db        462 TTTTTTTTNNNNTNNNNNNNNNNNNTTTTTTTTNTNTTTNNNNTNTNNNNNNNTTNNNTT 403

Qy        203 AGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATA 262
Db        402 TTTTTTTTTTTNNNNNNTNNNNNTTTTTNNTNNTNTTTTTTTTTTTTTNNNTTTTTTTTTN 343

Qy        263 ACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTC 322
Db        342 NNNNNTTTTTTNNNNNTTTTTTTTTNNTNNNNNTTTTTTTTTTNNNNNTTNTNTTTTTT 283

Qy        323 GAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCA 369
Db        282 NNNNNNNNNTTTNNTTTTTTTTTTTTTTTNTTNTTGGTTTTCAATACA 236
```

RESULT 10
CNS011ID/c
LOCUS CNS011ID 750 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN06D21 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL100303
VERSION AL100303.1 GI:5611914
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 750)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

FEATURES Location/Qualifiers
source 1. .750
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN06D21"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"

ORIGIN

Query Match 10.4%; Score 51.2; DB 29; Length 750;
Best Local Similarity 33.2%; Pred. No. 0.72;
Matches 113; Conservative 62; Mismatches 165; Indels 0; Gaps 0;

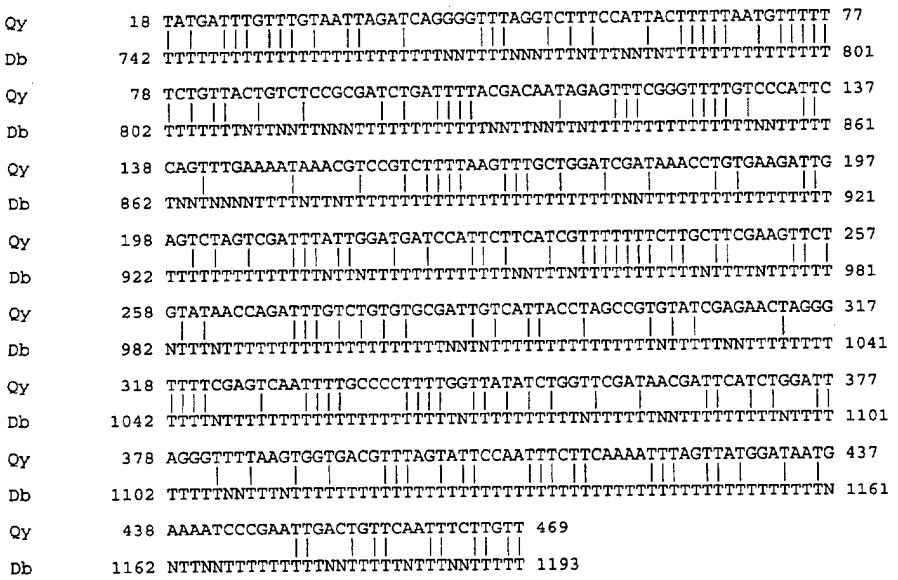
```
Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db        498 TKKTWTTKTTTTTWTTAWWTYTTWGKTTWGTTTTTTTWKKWATTTKTTTTTAAGWTTATK 439

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db        438 TTTGKTTTTTYKTTTTTTTTKTTTCAKTKTTGYTTTTKGATTTTKTTTTTKWTTYTTTTT 379

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db        378 TWKTTTTWTTWTCTAAGTTKTYTAGTTTTCTYTATTYCTTTAATKWTAKTTTTTTWTCTT 319

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db        318 KTAGYWTTWTTTGWTTTTKTTTTKWCTGTAKTKTKWKTTTYTCTWTTTTTKWKTAWTTTK 259

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db        258 KYTTAWTAKCTTTTKTTTTTKKWTKTTTTTTTATTAKTWTTCCTKTWTTKWKWTTTTWTA 199

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTC 357
Db        198 KTTTTCGAKKTTCTTTTTWTTYTTKKGTTWKYKTTTGTTY 159
```

RESULT 11
CG744146
LOCUS CG744146 1246 bp DNA linear GSS 24-OCT-2003
DEFINITION P036-3-E10.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG744146
VERSION CG744146.1 GI:37965014
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1246)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1246
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"

/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

ORIGIN

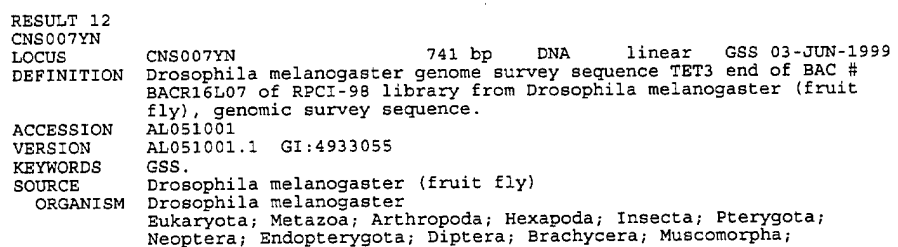
Query Match 10.3%; Score 51; DB 29; Length 1246;
Best Local Similarity 39.2%; Pred. No. 0.69;
Matches 177; Conservative 0; Mismatches 275; Indels 0; Gaps 0;



RESULT 12
CNS007YN
LOCUS CNS007YN 741 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR16L07 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL051001
VERSION AL051001.1 GI:4933055
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 741)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .741
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR16L07"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 10.1%; Score 50; DB 29; Length 741;
Best Local Similarity 34.6%; Pred. No. 1.3;
Matches 155; Conservative 0; Mismatches 293; Indels 0; Gaps 0;



RESULT 13
CNS007UZ
LOCUS CNS007UZ 563 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR16P14 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL050869
VERSION AL050869.1 GI:4932923
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 563)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

```
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR16P14"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             10.0%;  Score 49.6;  DB 29;  Length 563;
  Best Local Similarity   35.9%;  Pred. No. 1.7;
  Matches  148;  Conservative    0;  Mismatches  264;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db         74 TTTTTTTTTTTTGTTTTTTTTTTTTTTTTTTTGHCTTTTTTTTTTGTTTTTTTTTTTTT 133

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db        134 TTTTTTTTTTTTTTTTTTTTTTTNTNTTTTKKNNNNNNNTTTNTNNNTNTNNTTNNTTT 193

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db        194 NTNNTNNNNTTTTNTNNTNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 253

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db        254 TTTTTTTTTTTTTTNTTTTTTTNNNTTNNNTNNTNTTTTTTTCTTTNNCNNNTNNNTN 313

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db        314 NTNNATNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTNNTNNNTTNTTTTTNNTT 373

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db        374 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNTNTNTNNTNNTNTNNNTNTNTNNTTNTTT 433

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTAT 429
Db        434 NTNTTNTTTTTTNGTGNNGTNNNNNTNTTTTTTTTTNNNNNNTTTTTTTTTT 485


RESULT 14
CNS008BU
LOCUS       CNS008BU                737 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR16E03 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL051476
VERSION     AL051476.1  GI:4933530
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR16E03"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             10.0%;  Score 49.6;  DB 29;  Length 737;
  Best Local Similarity   33.3%;  Pred. No. 1.5;
  Matches  138;  Conservative   14;  Mismatches  263;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db        310 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 369

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db        370 NTTTTTNNTTTTTNNNTTNATTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTT 429

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db        430 TTNTTTTTTNTTTTTTTTTTNTCTNTTTTTTTTTTNCANTTGNNTWNTTTTTTWTTTTT 489

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db        490 TTTTTTNTNTTTTNTNANANNNTTMANNNTNTNCNNNTNTCTTTTTTNTTTTNTTNTTT 549

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db        550 TTTTNTTNTNTTTTNNNNNNNMNTTAWNNNNSNANCNANTNNNNGNANNNNANNNNNNN 609

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db        610 TWTTTTTTTTTTNTNYTTNTNTTTTTTTTTTTNTNWNTTTNNNHNTTNVNNTTNTTTTTNT 669

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGA 432
Db        670 NNSTTTTMNRAWNNTNATTTTTATTTNWTNNTTATTTNTNSATNNTTNTTWTTNA 724


RESULT 15
CG749401
LOCUS       CG749401                1403 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P043-3-E04.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG749401
VERSION     CG749401.1  GI:37970327
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
```

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

REFERENCE 1 (bases 1 to 1403)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007

COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .1403
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

ORIGIN

Query Match 10.0%; Score 49.6; DB 29; Length 1403;
Best Local Similarity 44.2%; Pred. No. 1.3;
Matches 196; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

```
Qy          27 TTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGTTACT 86
               ||||| ||   |      |||  | |||  ||  |||||  | ||||||  | ||  |
Db         362 TTTGTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 421

Qy          87 GTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTTTGAA 146
                | |      | |  ||||        |    |||   |||       ||   |||||
Db         422 TTTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGTGTTTTGTTGTTTGTT 481

Qy         147 AATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCTAGTC 206
               |     |  | ||||   |||  |   | |       | ||    ||   | | ||
Db         482 TGTTTTTTTTGTTGTTTTTGTTTTTTTTTTTTGTGGTTTTTNTGTTTTTTTTTTTTTGTT 541

Qy         207 GATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATAACCA 266
                ||| || | |  |   || ||  |  |||||| ||  ||      || | | |
Db         542 TTTTTTTTTGTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGT 601

Qy         267 GATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTCGAGT 326
               |||||  | | |    || |  ||   | |  || | | |     |    ||||    |
Db         602 TTTTTGTTTTTTTTTTTTTTTGTTTTTTTTGTTGTTTTTGGTTTGGTTTTTTTTTTTGTT 661

Qy         327 CAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTAGGGTTTTA 386
                ||||      |||| ||  | | |  || |     |  |  |  |  || | |||||
Db         662 TTTTTTTTGTGTTTTTTGTGTTTTTTTTTTTTGTGTTTGNGTTGTTGGTTTTTGTGTTTTT 721

Qy         387 AGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATGAAAATCCCG 446
                ||  |    ||| || ||     ||| ||      |||  || |   |  |      |
Db         722 GGTTTTTTTTTTTTGTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 781

Qy         447 AATTGACTGTTCAATTTCTTGTT 469
                ||   | ||   ||| || ||
Db         782 TTTTTTTTTTTTTTTTTTTTT 804
```

Search completed: June 20, 2004, 05:01:22
Job time : 2452.33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 23:55:13 ; Search time 366.167 Seconds
(without alignments)
5731.297 Million cell updates/sec

Title: US-09-000-062-7
Perfect score: 494
Sequence: 1 ctcaggcgaagaacaggtat.....cgcagatcccgggatctgcg 494

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	490.8	99.4	494	2	AAT85997	Aat85997 Arabidops
2	374.2	75.7	1005	3	AAC47929	Aac47929 Arabidops
3	84.6	17.1	677	3	AAC32694	Aac32694 Arabidops
4	83.6	16.9	674	3	AAC51182	Aac51182 Arabidops
5	55.8	11.3	12870	6	ABK39983	Abk39983 Human che
6	55.8	11.3	12870	6	ABL70229	Abl70229 Chemicall
7	49	9.9	17869	6	ABK39921	Abk39921 Human che
8	49	9.9	17869	6	ABL32105	Abl32105 Human imm
9	48	9.7	6593	6	ABL32478	Abl32478 Human imm
10	46.6	9.4	7814	4	AAS46530	Aas46530 Tumour su
11	46.4	9.4	6650	6	ABL32638	Abl32638 Human imm
12	46.2	9.4	6048	6	ABL32509	Abl32509 Human imm
13	46	9.3	3025	7	ABZ10189	Abz10189 Haematopo
14	46	9.3	3025	9	ADB54219	Adb54219 Pretreate
15	46	9.3	3025	9	ADE84157	Ade84157 Human lym
16	45.6	9.2	6668	6	ABL33697	Abl33697 Human imm
17	45.6	9.2	7442	4	AAS46686	Aas46686 Tumour su
18	44.8	9.1	5208	6	ABL32923	Abl32923 Human imm
19	44.8	9.1	12733	6	ABL32091	Abl32091 Human imm
20	44.4	9.0	9547	6	ABL33505	Abl33505 Human imm
21	44.2	8.9	594	6	ABQ46990	Abq46990 Oligonucl
c 22	44.2	8.9	594	6	ABQ46991	Abq46991 Oligonucl
23	44	8.9	5204	6	ABL32899	Abl32899 Human imm
24	44	8.9	5989	6	ABL54320	Abl54320 Chemicall
25	44	8.9	18585	6	ABL34609	Abl34609 Human met
26	43.8	8.9	6467	6	ABN80169	Abn80169 Human che
27	43.6	8.8	18598	6	ABL32387	Abl32387 Human imm
28	43.4	8.8	3025	7	ABZ10043	Abz10043 Haematopo
29	43.4	8.8	3025	9	ADB54091	Adb54091 Pretreate
30	43.4	8.8	3025	9	ADE84081	Ade84081 Human lym
31	43.4	8.8	6523	9	ADE84216	Ade84216 Human lym
32	43.4	8.8	16170	6	ABL33269	Abl33269 Human imm
33	43.2	8.7	3683	7	ABZ10199	Abz10199 Haematopo
34	43.2	8.7	5394	6	ABL33008	Abl33008 Human imm
35	43.2	8.7	6668	6	ABL33696	Abl33696 Human imm
36	43.2	8.7	9964	6	ABL32098	Abl32098 Human imm
37	43.2	8.7	12705	6	ABL32148	Abl32148 Human imm
38	42.8	8.7	3001	6	ABN80333	Abn80333 Human che
39	42.8	8.7	6120	6	ABK28439	Abk28439 DNA trans
40	42.8	8.7	6120	6	ABN80302	Abn80302 Human che
41	42.6	8.6	866	4	AAI94068	Aai94068 Human neu
42	42.6	8.6	5552	6	ABL33258	Abl33258 Human imm
43	42.6	8.6	17918	6	AAS61418	Aas61418 Human gen
44	42.4	8.6	7657	4	AAS45477	Aas45477 Chemicall
45	42.4	8.6	7657	6	ABL34022	Abl34022 Human imm

ALIGNMENTS

RESULT 1
AAT85997
ID AAT85997 standard; DNA; 494 BP.
XX
AC AAT85997;
XX
DT 17-NOV-1997 (first entry)
XX
DE Arabidopsis thaliana histone H3.3-like DNA fragment (intron 2).
XX
KW Plant expression regulation sequence; intron 2; histone;
KW herbicide tolerance; 5-enolpyruvylshikimate-3-phosphate synthase; EPSPS;
KW glyphosate; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO9704114-A2.
XX
PD 06-FEB-1997.
XX
PF 17-JUL-1996; 96WO-FR001109.
XX
PR 19-JUL-1995; 95FR-00008980.
XX
PA (RHON) RHONE POULENC AGROCHIMIE.
XX
PI Derose R, Chaubet N, Gigot C;
XX
DR WPI; 1997-132652/12.
XX
PT New regulatory sequence for chimeric gene expression in rapidly growing
PT parts of a plant - includes at least one intron from a plant histone gene
PT and is useful for imparting resistance to herbicides.
XX
PS Claim 4; Page 27; 31pp; French.
XX
CC The known cosmid clone c22 of Arabidopsis thaliana contains two histone
CC H3.3-like genes. Digestion of clone c22 with restriction enzymes AluI and
CC CfoI generated a fragment of 494 bp having the present sequence. This
CC fragment, designated intron 2, was ligated to synthetic linkers for
CC cloning into plant expression vectors. In addition to the intron 2
CC sequence, the vectors contained a plant promoter and a herbicide
CC tolerance gene (e.g. a mutated version of the 5-enolpyruvylshikimate-3-
CC phosphate synthase (EPSPS) gene). The intron enhances expression of the
CC herbicide tolerance gene in rapidly growing parts of plants. The intron
CC can also be used to enhance expression of genes that impart resistance to

```
CC   pathogens or that encode nutritional or therapeutic proteins
XX
SQ   Sequence 494 BP; 113 A; 83 C; 104 G; 194 T; 0 U; 0 Other;

  Query Match             99.4%;  Score 490.8;  DB 2;  Length 494;
  Best Local Similarity   99.6%;  Pred. No. 5.5e-118;
  Matches  492;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 180

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        421 TTTAGTTATGGATAATGAAAATCCCCAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480

Qy        481 TCCCGGGATCTGCG 494
              ||||||||||||||
Db        481 TCCCGGGATCTGCG 494


RESULT 2
AAC47929
ID   AAC47929 standard; DNA; 1005 BP.
XX
AC   AAC47929;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 55633.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic pathway;
KW   promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134221P.
PR   14-MAY-1999;   99US-0134370P.
PR   18-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137528P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   18-JUN-1999;   99US-0139750P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   22-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   23-JUN-1999;   99US-0140354P.
PR   24-JUN-1999;   99US-0140695P.
PR   28-JUN-1999;   99US-0140823P.
PR   29-JUN-1999;   99US-0140991P.
PR   30-JUN-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   01-JUL-1999;   99US-0142154P.
PR   02-JUL-1999;   99US-0142055P.
PR   06-JUL-1999;   99US-0142390P.
PR   08-JUL-1999;   99US-0142803P.
PR   09-JUL-1999;   99US-0142920P.
PR   12-JUL-1999;   99US-0142977P.
PR   13-JUL-1999;   99US-0143542P.
PR   14-JUL-1999;   99US-0143624P.
PR   15-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144085P.
PR   16-JUL-1999;   99US-0144086P.
PR   19-JUL-1999;   99US-0144325P.
```

PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 75.7%; Score 374.2; DB 3; Length 1005;
Best Local Similarity 95.5%; Pred. No. 1.5e-87;
Matches 385; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy         19 ATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTT 78
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTT 60

Qy         79 CTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCC 138
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCC 120

Qy        139 AGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGA 198
               |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGA 180

Qy        199 GTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTG 258
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTG 240

Qy        259 TATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGT 318
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGT 300

Qy        319 TTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTA 378
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTA 360

Qy        379 GGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAAT 421
               ||||||||||||| || |   ||  |  | ||  ||  || |
Db        361 GGGTTTTAAGTGATGGCTCGTACCAAGCAAACCGCTCGTAAGT 403
```

RESULT 3
AAC32694
ID AAC32694 standard; DNA; 677 BP.
XX
AC AAC32694;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 305.
XX
KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.

PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 17.1%; Score 84.6; DB 3; Length 677;
Best Local Similarity 77.9%; Pred. No. 3.1e-12;
Matches 102; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy         69 AATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTT 128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTT 60

Qy        129 GTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTG 188
              |||||||||||||||||||||  |   || |  |      | |||  |  ||
Db         61 GTCCCATTCCAGTTTGAAAATAAAGATGGCTCGTACCAAGCAAACCGCTCGTAAGTCCAC 120

Qy        189 TGAAGATTGAG 199
              | || |  ||
Db        121 CGGAGGTAAAG 131
```

RESULT 4
AAC51182
ID AAC51182 standard; DNA; 674 BP.
XX
AC AAC51182;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67569.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 16.9%; Score 83.6; DB 3; Length 674;
Best Local Similarity 77.7%; Pred. No. 5.6e-12;
Matches 101; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy         70 ATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTG 129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTG 60

Qy        130 TCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGT 189
              ||||||||||||||||||||  |  || |  |     | |||  |  ||
Db         61 TCCCATTCCAGTTTGAAAATAAAGATGGCTCGTACCAAGCAAACCGCTCGTAAGTCCACC 120

Qy        190 GAAGATTGAG 199
              | ||| ||
Db        121 GGAGGTAAAG 130
```

RESULT 5
ABK39983
ID ABK39983 standard; DNA; 12870 BP.
XX
AC ABK39983;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #33 strand 1.

```
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 65; 24pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence at least 18
CC   bases in length of a segment of the chemically pretreated DNA of genes
CC   associated with pharmacogenomics according to one of the sequences of the
CC   genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC   (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC   (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC   NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC   their complementary sequences, or a sequence (S1) chosen from 87
CC   sequences and their complements. The chemical pretreatment is bisulphite
CC   treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC   Also included are an oligomer (II) in particular an oligonucleotide or a
CC   peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC   base sequence having a length of 9 nucleotides which hybridises to or is
CC   identical to a chemically pretreated DNA of genes associated with
CC   pharmacogenomics and their complements, arranged in an array for
CC   analysing diseases associated with the methylation state (CpG) and/or
CC   detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC   oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC   and their complements is useful for diagnosis and therapy of solid
CC   tumours and cancer. The present sequence represents one the 87 DNA
CC   sequences or its complement. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 12870 BP; 3115 A; 234 C; 3273 G; 6248 T; 0 U; 0 Other;

  Query Match           11.3%;  Score 55.8;  DB 6;  Length 12870;
  Best Local Similarity   46.5%;  Pred. No. 0.0002;
  Matches  180;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
              | |  || ||||  |||    |     |||| || |||  |||  ||  || | |||||
Db       3147 TTTATTTGGTTTTAAATGTTTTTTATATTTTAAGTTTTTTAATTGATTGATAGTTTTTTT 3206

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
              | |  |  | |      | | || ||||      |          |||||||  |  ||
Db       3207 TATAGTTTTTTAATATAGTTTTGTTTTTTTATTGGTTTTTAGATATGTTTTGTTCGTTTT 3266

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
              | ||||     |  |  |  |||   |||  ||||  | |||  | ||      ||
Db       3267 TATTTTGTGTTTTTATTTATATTGTTTTTTTTTTTTGGAATGTTAATTCGGTTTTTTTTT 3326

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
                |    |   ||  |   | ||   ||  |     |||| || ||  |  |||||| |

Db       3327 TTTTAGTTTTTATTTGTTTTTTATTTTTTTTATTTAAATTTTATTTTTTTTTAGGAAGTTTT 3386

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
              | | |  |  || || | | |    |  |  ||   | |   | | |   ||| ||  |
Db       3387 TTTTGATTATTTTAGTTTATTTTTATTATTTTTTTTTTTTGGATTTTTTTAGGAATTAAAG 3446

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
              ||   || |   | |||     |||| | |||  |  || | |   | ||  | |   ||
Db       3447 TTAGTGAATACGTATTGATATTTTTTATAATAATTTATTGGGTTTTGGTTTTTTTTTTTT 3506

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTAT 404
                |   ||  || | |     |||| |
Db       3507 TAGAAGTTCGGTAGGGTATGATAGTTT 3533


RESULT 6
ABL70229
ID   ABL70229 standard; DNA; 12870 BP.
XX
AC   ABL70229;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated cell signalling DNA sequence#60.
XX
KW   Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW   tumour; cytostatic; ds.
XX
OS   Unidentified.
XX
PN   WO200202807-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007471.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
XX
PT   Nucleic acid, useful for diagnosis and therapy of diseases associated
PT   with cell signaling e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with cell signaling.
XX
PS   Claim 1; SEQ ID NO 119; 24pp + Sequence Listing; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of at least
CC   18 bases of a segment of chemically pretreated DNA of genes associated
CC   with cell signalling. The activity of the modified sequences of the
CC   invention may be described as cytostatic. The object of the invention is
CC   to provide the chemically modified DNA of genes associated with cell
CC   signalling, as well as oligonucleotides and/or PNA-oligomers for
CC   detecting cytosine methylations, as well as a method which is
CC   particularly suitable for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with cell signalling. The
CC   chemically modified DNA provided by the invention is useful for diagnosis
CC   and therapy of diseases such as solid tumours and cancer. The sequences
CC   given in records ABL70111-ABL70626 represent chemically pre-treated
CC   genomic DNA's of genes associated with cell signalling. Note: The
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office
XX
SQ   Sequence 12870 BP; 3115 A; 234 C; 3273 G; 6248 T; 0 U; 0 Other;

  Query Match           11.3%;  Score 55.8;  DB 6;  Length 12870;
  Best Local Similarity   46.5%;  Pred. No. 0.0002;
  Matches  180;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;
```

```
Qy          18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
               | |  || ||||  |||     |       |||| || |||  |||  ||  || | |||||
Db        3147 TTTATTTGGTTTTAAATGTTTTTTATATTTTAAGTTTTTTAATTGATTGATAGTTTTTT 3206

Qy          78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
               | |   | | |      | | || ||||        |          |||||||  |  ||
Db        3207 TATAGTTTTTTAATATAGTTTTGTTTTTTTATTGGTTTTTAGATATGTTTTGTTCGTTTT 3266

Qy         138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
               | ||||     |  |   |  |||   |||  ||||  | |||  | ||      ||
Db        3267 TATTTTGTGTTTTTATTTATATTGTTTTTTTTTTTTGGAATGTTAATTCGGTTTTTTTTT 3326

Qy         198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
               |    |    ||  |   | ||   ||  |    ||||  || ||  |  ||||||  |
Db        3327 TTTTAGTTTTTATTTGTTTTTTATTTTTTTATTTAAATTTTATTTTTTTTTAGGAAGTTTT 3386

Qy         258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
               | | |  |  || || | |  |  |  ||  | |  | |  |  ||| ||  |
Db        3387 TTTTGATTATTTTAGTTTATTTTTATTATTTTTTTTTTTGGATTTTTTTAGGAATTAAAG 3446

Qy         318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
               ||   || |  | |||    ||||  | |||  |  || | |  | || | |   ||
Db        3447 TTAGTGAATACGTATTGATATTTTTTATAATAATTTATTGGGTTTTGGTTTTTTTTTTTT 3506

Qy         378 AGGGTTTTAAGTGGTGACGTTTAGTAT 404
               |   ||  || | |    |||| |
Db        3507 TAGAAGTTCGGTAGGGTATGATAGTTT 3533
```

RESULT 7
ABK39921
ID ABK39921 standard; DNA; 17869 BP.
XX
AC ABK39921;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #1 strand 2.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 2; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 17869 BP; 5366 A; 158 C; 3365 G; 8978 T; 0 U; 2 Other;

  Query Match   9.9%; Score 49; DB 6; Length 17869;
  Best Local Similarity   46.0%; Pred. No. 0.012;
  Matches 210; Conservative   0; Mismatches 240; Indels   7; Gaps   1;

```
Qy          18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
               | || |||  || |  | | |         ||| | |||   ||  ||| |  | |||||
Db     17051 TTTGTTTTTGTTTTTTTGAAAATTTATTGATAGTTTTTTATTTTTTTTTGTTGTTTTTTT 17110

Qy          78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
               |   ||  |   |   || | |   ||    | |  | || ||| |  |  |     |||
Db     17111 TTGTTTGTTTATTTTTCGTTTTTTTTGGTTGTCGGTTGATTTTAGTTTGGTTGTGTATTT 17170

Qy         138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
                ||||    |      ||    |||| | |||| | |      ||      ||  || ||
Db     17171 GTTTTTGGTTGTTTTTTTCGTGTTTTTGATTTTGTTTGTGTTTTATTTAGGTTTAGTTTT 17230

Qy         198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
               ||          ||| || | | ||   |   |     ||| ||| ||  |||| | || |
Db     17231 TGT-------TTTTTTTTGGTTATTTTTGTGTGTGTTTTTGTTTTTTTTGTTCGTATTTTT 17283

Qy         258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
               |  |          || || ||| |    || |  ||   ||  ||  ||||        |
Db     17284 GGTTGTTTTTGTTAGTTTGTTTTTATTTTTATTTGTTTATTTGTTTGAAGAGATAGGTAG 17343

Qy         318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
               ||||   ||   |||        ||||  | | | || ||   | |   ||  | ||| ||
Db     17344 TTTTTTTGTTTTTTTATATTTATTTGTGTTTTTTTGTTTTTGTGATTTTTTTTTTGGTTT 17403

Qy         378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATG 437
                  ||    |  ||   ||| |||||     ||| ||        ||  || |   | ||
Db     17404 TTTTGTTGTTTTTTTGTTTTTTTGTATTTTTGTTTGTTTTTTTTGTTTATTGTTATTTATT 17463

Qy         438 AAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATG 474
                  |   || ||   | || | ||| || ||   ||
Db     17464 TTTTTTTAGAGTTTGTTTTTATTTTTTTTTTTTTTG 17500
```

RESULT 8
ABL32105
ID ABL32105 standard; DNA; 17869 BP.
XX
AC ABL32105;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 78.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

```
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 78; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 17869 BP; 5366 A; 158 C; 3365 G; 8978 T; 0 U; 2 Other;

  Query Match             9.9%;  Score 49;  DB 6;  Length 17869;
  Best Local Similarity   46.0%;  Pred. No. 0.012;
  Matches  210;  Conservative    0;  Mismatches  240;  Indels    7;  Gaps    1;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db      17051 TTTGTTTTTGTTTTTTTGAAAATTTATTGATAGTTTTTATTTTTTTTTGTTGTTTTTTT 17110

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db      17111 TTGTTTGTTTATTTTTCGTTTTTTTTGGTTGTCGGTTGATTTTAGTTTGGTTGTGTATTT 17170

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db      17171 GTTTTTGGTTGTTTTTTTCGTGTTTTTGATTTTGTTTGTGTTTTATTTAGGTTTAGTTTT 17230

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db      17231 TGT-------TTTTTTTTGGTTATTTTTGTGTGTGTTTTTGTTTTTTTGTTCGTATTTTT 17283

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db      17284 GGTTGTTTTTGTTAGTTTGTTTTTATTTTTATTTGTTTATTTGTTTGAAGAGATAGGTAG 17343

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db      17344 TTTTTTTGTTTTTTTATATTTATTTGTGTTTTTTTGTTTTTGTGATTTTTTTTTTGGTTT 17403

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATG 437
Db      17404 TTTTGTTGTTTTTTTTGTTTTTTTGTATTTTTGTTTGTTTTTTTGTTTATTGTTATTTATT 17463

Qy        438 AAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATG 474
Db      17464 TTTTTTTAGAGTTTTGTTTTTATTTTTTTTTTTTTTG 17500
```

```
RESULT 9
ABL32478
ID   ABL32478 standard; DNA; 6593 BP.
XX
AC   ABL32478;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 451.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 451; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6593 BP; 1732 A; 121 C; 1448 G; 3292 T; 0 U; 0 Other;

  Query Match             9.7%;  Score 48;  DB 6;  Length 6593;
  Best Local Similarity   45.8%;  Pred. No. 0.018;
  Matches  207;  Conservative    0;  Mismatches  240;  Indels    5;  Gaps    1;

Qy         23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
Db        274 TTTTTTTTTATTTATTGTATGTATTTTTTTTATGGTTTTTATGTTTTTTTGTTTTTTTTT 333

Qy         83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
Db        334 TTTTTTTTTGGATATTTATAATTAATATATTTGTTTATTTGGTGGTGTTTGATAATTTTT 393

Qy        143 TGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGT-----GAAGATTG 197
Db        394 TTAGATTTTTTTTATTTTTTTTGATTTTTTTTTGTTTTTTGTTTTTTTTGATTGGATAATTT 453

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
```

```
Db        454 TAAATGATTTGTTTTGGAGTTGATTAATTTTTTTGTTTGATTAAATTTGTTGTTAAATTT 513

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317

Db        514 TTATGATGAATTTTTTATTTTAGTTATTGTATTTTTTTATTTTAGAATTTTTATTTGGTT 573

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377

Db        574 TTTTTTTATTTTTTTTATTTTTTTATAATATTTTTATTTATTTAATGTGTTATTTTTTTT 633

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATG 437

Db        634 ATTTTGTTTAGTTGTTTATTTGGGTTTTTTTGTAATTTGTTGAGTTATTTTAAGATGATT 693

Qy        438 AAAATCCCGAATTGACTGTTCAATTTCTTGTT 469

Db        694 ATTTTGAATTTTTTGTTAGATAATTTATAGAT 725
```

RESULT 10
AAS46530
ID AAS46530 standard; DNA; 7814 BP.
XX
AC AAS46530;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #252.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 252; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7814 BP; 1677 A; 101 C; 1779 G; 4257 T; 0 U; 0 Other;

Query Match 9.4%; Score 46.6; DB 4; Length 7814;
Best Local Similarity 46.8%; Pred. No. 0.043;
Matches 214; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

```
Qy         21 GATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCT 80

Db       3356 GGTTTTTTTATTATTAGAAGAGGGTTTTGGGTTTTTATATTTTTTTTTTTTTAGTTTTTTG 3415

Qy         81 GTTACTGTCTCCGCGATCTGA-TTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCA 139

Db       3416 ATTATTATATTTAAATTTTTAGGGTTTTGTTTTTAGTTTAGGTTTTTTAGTTTTTTTATT 3475

Qy        140 GTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAG 199

Db       3476 TTTTGGTAATGTTTGAATATTAATTTAATTTTTTTGGTTTATAAATTTTTTTTGTTTATT 3535

Qy        200 TCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGT 259

Db       3536 TATTATTTATTTGTTTTTTTATGTATTTATTTGTTTGTTTATTTGTTTATTTAGTTATTT 3595

Qy        260 ATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTT 319

Db       3596 ATTTATTAAGTTTATTATGTATTTATTTATGTGTTTATTTATTTAATTGTTCGTTTGTTT 3655

Qy        320 TTCGAGTCAATTTTGCCCCTTTTGGTTA---TATCTGGTTCGATAACGATTCATCTGGAT 376

Db       3656 GTTTATTTATTTGTTAGTTAATTTATTTATTTATTTATTTGTTTATTTATTTGTTTATTT 3715

Qy        377 TAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAAT 436

Db       3716 AATTATTTATATGTTTATTAATTAGTTCGTTTATTTGTTATTAGTTTATTAGTTTATAAT 3775

Qy        437 GAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAAT 473

Db       3776 TTTATTTGTTTATTTATTTATGAATTTGTTTATTTAT 3812
```

RESULT 11
ABL32638
ID ABL32638 standard; DNA; 6650 BP.
XX
AC ABL32638;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 611.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX

PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 611; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6650 BP; 1628 A; 159 C; 1561 G; 3302 T; 0 U; 0 Other;

Query Match 9.4%; Score 46.4; DB 6; Length 6650;
Best Local Similarity 44.0%; Pred. No. 0.047;
Matches 197; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

```
Qy        23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
Db      3598 TGTGTTTTATTTTTTTTATTTTGTTTTTTTTTTTATTTTTATTTATGTTTTTTTTTTTT 3657

Qy        83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
Db      3658 TTATTAGTTTAGGTTTTAGGTTTATATATAAGTGTTTTCGGTATTGTTGAGTTTTTATGA 3717

Qy       143 TGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCT 202
Db      3718 TGATATTTAAGATGTTTGTTGTAAGTTAATAAAATTGTTAATTATTTATTATTTAATAAT 3777

Qy       203 AGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATA 262
Db      3778 ACGGAATGAGAGTGCGTAGTGATTATGAAGTATATATGGTTTTTTTGTTATGGTTATTTC 3837

Qy       263 ACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTC 322
Db      3838 GTTTTTTATAATTTTAGATTAGTAATTTTTGTTAATTGTGATTTTTGATTTAATTTTTTT 3897

Qy       323 GAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATTAGGGT 382
Db      3898 TATTTTTTATTTTAGGGTTTATTGTTTTGTTTTTTGTTATTTTTTTTTTTTTATTTGAAT 3957

Qy       383 TTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATGAAAAT 442
Db      3958 TCGTTTTTGATTGGTGTAGTTGTTGAAGATGTTGTAGTTGTAGTAGTGTATATTTGGAAT 4017

Qy       443 CCCGAATTGACTGTTCAATTTCTTGTTA 470
Db      4018 ATTTGGTTGATTTGTTGAATTTTTTTTA 4045
```

RESULT 12
ABL32509
ID ABL32509 standard; DNA; 6048 BP.
XX
AC ABL32509;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 482.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 482; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6048 BP; 1296 A; 103 C; 1361 G; 3288 T; 0 U; 0 Other;

Query Match 9.4%; Score 46.2; DB 6; Length 6048;
Best Local Similarity 46.9%; Pred. No. 0.051;
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

```
Qy       163 TTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATC 222
Db      2325 TTAAAGTTTTTTGGATGGTTTATAAGGTTTTATAAAATTTGGTATTTTTGTTATTTTTTG 2384

Qy       223 CATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCG 282
Db      2385 GATTTTTTTTTTGTTATTTTTTTTTTGTTGTTTTGTTTTAGTTATTTTGTGTTTTTTGT 2444

Qy       283 ATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTT 342
Db      2445 TTTTTGAATTTGTTATGTTGTTTTTTGTTTTAGGGTTTTTGTATTTGTTGTTTTATTTGT 2504

Qy       343 GGTTATATCTGGTTCGATAACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGT 402
Db      2505 TTGGATTTTTTTTTTCGTAAATAGTTATATGGGTTGGTTTTTTTTTTTTTTAGGTTTTT 2564

Qy       403 ATTCCAATTTCTTCAAAATTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATT 462
Db      2565 GTTTAAATGTTATTTTTTTATTGAGGTTTTTTTTGAGTATTTTTTTTAAAATTTGTAAGT 2624

Qy       463 TCTTGTT 469
```

```
Db       2625 TTTTTTT 2631


RESULT 13
ABZ10189
ID   ABZ10189 standard; DNA; 3025 BP.
XX
AC   ABZ10189;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #329.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 329; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 3025 BP; 535 A; 0 C; 930 G; 1560 T; 0 U; 0 Other;

  Query Match             9.3%;  Score 46;  DB 7;  Length 3025;
  Best Local Similarity   47.9%;  Pred. No. 0.049;
  Matches  162;  Conservative    0;  Mismatches  175;  Indels    1;  Gaps    1;

Qy         23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
              ||  ||||   ||   | |||   ||     |||  |||  | |   | | |||| |

Db       1881 TTGTTTTGGTTTTTATTTAGGATGTTTTTATTTTTATTTATATATGTTTATGTTTTGGGA 1940

Qy         83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
                || |  | |||   || ||| |  | |||  ||| |  ||||       ||   | |

Db       1941 AGTGGTGTAGGAGATGAAATATTAAGTAAGTAGTTTTTTGTTTTTTTGATTGTTTTGGAT 2000

Qy         143 TGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCT 202
               |  || ||||  |      ||||  | ||||  |  |  |  ||       |   |  || |

Db       2001 TTTAATTAAAGTTTTTAGTTTTTAATTTGTTTTAAAG-TATTGGGTTTGGGGGTGGGAGG 2059

Qy         203 AGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATA 262
                 |     |  ||  ||  |   || || || |  ||||| |||  | |||  ||  || |

Db       2060 TTTGTTGTGGTTTTATTTTTGTTTATTAATTGAGTTTTTTTTTGTGTAGAATTTAGGTTTG 2119

Qy         263 ACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTC 322
                 ||| |   ||      || | |   |  | |    |||| | | | |   ||||

Db       2120 GGAAGAGTGTGGTGGTGTTGAGTTTAGATTGTTTGTTTAGTATTGTGGAGTGTATTTTTA 2179

Qy         323 GAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
                | | |||||     |||| || | | | | |||

Db       2180 TTGAGAGTTTTGTGGTGTTTGTTTTTTTGTTGGTGGAT 2217


RESULT 14
ADB54219
ID   ADB54219 standard; DNA; 3025 BP.
XX
AC   ADB54219;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Pretreated genomic DNA region 143.
XX
KW   colon cell proliferative disorder; non methylated CpG dinucleotide;
KW   cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS   Unidentified.
XX
PN   WO2003072821-A2.
XX
PD   04-SEP-2003.
XX
PF   27-FEB-2003; 2003WO-EP002035.
XX
PR   27-FEB-2002; 2002EP-00004551.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Adorjan P,  Burger M,  Maier S,  Nimmrich I,  Becker E,  Lesche R;
PI   Rujan T,  Schmitt A;
XX
DR   WPI; 2003-731620/69.
XX
PT   Detecting and differentiating between colon cell proliferative disorders
PT   associated with a gene or its regulatory regions comprises contacting a
PT   target nucleic acid in a biological sample obtained from the subject with
PT   a reagent.
XX
PS   Claim 32; SEQ ID NO 275; 74pp; English.
XX
CC   The invention relates to a novel method for detecting and differentiating
CC   between colon cell proliferative disorders associated with at least one
CC   gene or its regulatory regions. The method comprises contacting a target
CC   nucleic acid in a biological sample obtained from the subject with at
CC   least one reagent or a series of reagents, where the reagent or series of
CC   reagents, distinguishes between methylated and non methylated CpG
CC   dinucleotides within the target nucleic acid. The molecules of the
CC   invention demonstrate cytostatic activity whilst the method may useful
CC   for detecting and differentiating between colon cell proliferative
CC   disorders, including cancers such as colon adenoma and colon carcinoma.
CC   The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC   determining cytosine methylation state or single nucleotide
CC   polymorphisms. The current sequence is that of the pretreated genomic DNA
```

CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 3025 BP; 535 A; 0 C; 930 G; 1560 T; 0 U; 0 Other;

Query Match 9.3%; Score 46; DB 9; Length 3025;
Best Local Similarity 47.9%; Pred. No. 0.049;
Matches 162; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

```
Qy         23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
Db       1881 TTGTTTTGGTTTTTATTTAGGATGTTTTTATTTTTATTTATATATGTTTATGTTTTGGGA 1940

Qy         83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
Db       1941 AGTGGTGTAGGAGATGAAATATTAAGTAAGTAGTTTTTTGTTTTTTTGATTGTTTTGGAT 2000

Qy        143 TGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCT 202
Db       2001 TTTAATTAAAGTTTTTAGTTTTTAATTTGTTTTAAAG-TATTGGGTTTGGGGGTGGGAGG 2059

Qy        203 AGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATA 262
Db       2060 TTTGTTGTGGTTTTATTTTTGTTTATTAATTGAGTTTTTTTTTGTGTAGAATTTAGGTTTG 2119

Qy        263 ACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTC 322
Db       2120 GGAAGAGTGTGGTGGTGTTGAGTTTAGATTGTTTGTTTAGTATTGTGGAGTGTATTTTTA 2179

Qy        323 GAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
Db       2180 TTGAGAGTTTTGTGGTGTTTGTTTTTTTGTTGGTGGAT 2217
```

RESULT 15
ADE84157
ID ADE84157 standard; DNA; 3025 BP.
XX
AC ADE84157;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human lymphoid cell proliferative disorder gene derived DNA #93.
XX
KW ds; lymphoid cell proliferative disorder; methylation;
KW methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KW diffuse large B-cell lymphoma; mantle cell lymphoma;
KW chronic lymphocytic leukemia; small lymphocytic lymphoma;
KW follicular lymphoma; diagnosis; prognosis.
XX
OS Homo sapiens.
XX
PN WO2003044226-A2.
XX
PD 30-MAY-2003.
XX
PF 25-NOV-2002; 2002WO-EP013265.
XX
PR 23-NOV-2001; 2001DE-01057491.
PR 28-DEC-2001; 2001DE-01064501.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimmrich I;
XX
DR WPI; 2003-457621/43.
XX
PT Detecting and differentiating between lymphoid cell proliferative
PT disorders comprises contacting a target nucleic acid with at least one
PT reagent that distinguishes between methylated and non-methylated CpG
PT dinucleotides.
XX
PS Claim 26; SEQ ID NO 153; 448pp; English.

XX
CC The invention relates to a method of detecting and differentiating
CC between lymphoid cell proliferative disorders associated with at least
CC one gene and/or their regulatory regions in a subject by contacting a
CC target nucleic acid in a biological sample obtained from the subject with
CC at least one reagent or series of reagents that distinguish between
CC methylated and non-methylated CpG dinucleotides within the target nucleic
CC acid. The genes and/or their regulatory regions are preferably selected
CC from MDR1, CSNX2B, EGR4, AR, CDK4, RB2, CDC25A, GPIb beta, MYOD1, CDH3,
CC MYCL1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2a, CDKN2B, FOS,
CC GSTP1, HIC-1, MGMT, MLH1, MOS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKNIC,
CC GSK3beta, ESR1, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
CC acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
CC of the genes are useful for detecting the methylation state of all the
CC CpG dinucleotides within one or more the sequences, or their complements,
CC for determining the cytosine methylation state and or single nucleotide
CC polymorphisms (SNPs), and for differentiating at least two of the medical
CC conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,
CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular
CC lymphoma. They are also useful for detecting of a predisposition to,
CC differentiation between subclasses, diagnosis, prognosis, treating and/or
CC monitoring of lymphoid cell proliferative disorder. This sequence
CC represents a nucleic acid of a pretreated genomic DNA derived from the
CC above mentioned genes.
XX
SQ Sequence 3025 BP; 535 A; 0 C; 930 G; 1560 T; 0 U; 0 Other;

Query Match 9.3%; Score 46; DB 9; Length 3025;
Best Local Similarity 47.9%; Pred. No. 0.049;
Matches 162; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

```
Qy         23 TTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTCTGT 82
Db       1881 TTGTTTTGGTTTTTATTTAGGATGTTTTTATTTTTATTTATATATGTTTATGTTTTGGGA 1940

Qy         83 TACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCAGTT 142
Db       1941 AGTGGTGTAGGAGATGAAATATTAAGTAAGTAGTTTTTTGTTTTTTTGATTGTTTTGGAT 2000

Qy        143 TGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAGTCT 202
Db       2001 TTTAATTAAAGTTTTTAGTTTTTAATTTGTTTTAAAG-TATTGGGTTTGGGGGTGGGAGG 2059

Qy        203 AGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATA 262
Db       2060 TTTGTTGTGGTTTTATTTTTGTTTATTAATTGAGTTTTTTTTTGTGTAGAATTTAGGTTTG 2119

Qy        263 ACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTC 322
Db       2120 GGAAGAGTGTGGTGGTGTTGAGTTTAGATTGTTTGTTTAGTATTGTGGAGTGTATTTTTA 2179

Qy        323 GAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
Db       2180 TTGAGAGTTTTGTGGTGTTTGTTTTTTTGTTGGTGGAT 2217
```

Search completed: June 20, 2004, 01:58:58
Job time : 368.167 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 00:55:23 ; Search time 3463.42 Seconds
(without alignments)
6182.178 Million cell updates/sec

Title: US-09-000-062-7
Perfect score: 494
Sequence: 1 ctcaggcgaagaacaggtat.....cgcagatcccgggatctgcg 494

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		494	100.0	494	6	AR182672	AR182672 Sequence
2		490.8	99.4	494	6	A59350	A59350 Sequence 7
3		477.8	96.7	4833	8	ATH3G	X60429 A.thaliana
4	c	477.8	96.7	112067	8	ATCHRIV92	AL161596 Arabidops
5	c	477.8	96.7	118267	8	ATT5J17	AL035708 Arabidops
6		476.2	96.4	777	8	AJ592632	AJ592632 Arabidops
7		55.8	11.3	12870	6	AX348370	AX348370 Sequence
8		55.8	11.3	12870	6	AX348661	AX348661 Sequence
9	c	50.6	10.2	14281	8	SKORFS	X54850 S.kluyveri
10		49.6	10.0	7218	6	I66494	I66494 Sequence 14
11		49.6	10.0	349980	6	AX344573	AX344573 Sequence
12		49	9.9	17869	6	AX345007	AX345007 Sequence
13		49	9.9	17869	6	AX348307	AX348307 Sequence
14	c	49	9.9	176162	5	AL953898	AL953898 Zebrafish
15		48.6	9.8	250029	3	AE014820	AE014820 Plasmodiu
16		48	9.7	6593	6	AX345380	AX345380 Sequence
17		47.8	9.7	93791	2	AC138073	AC138073 Homo sapi
18	c	47.2	9.6	165888	2	AC144291	AC144291 Macaca mu
19	c	47.2	9.6	166224	2	AC142946	AC142946 Macaca mu
20		47	9.5	1453	8	AJ591978	AJ591978 Arabidops
21		47	9.5	95764	2	AC015732	AC015732 Homo sapi
22		46.6	9.4	7814	6	AX251284	AX251284 Sequence
23		46.4	9.4	6650	6	AX345540	AX345540 Sequence
24		46.4	9.4	37159	2	AC100259	AC100259 Mus muscu
25		46.4	9.4	56693	2	AC084077	AC084077 Homo sapi
26		46.2	9.4	6048	6	AX345411	AX345411 Sequence
27	c	46.2	9.4	164103	2	AC037484	AC037484 Homo sapi
28	c	46.2	9.4	190701	9	AC022274	AC022274 Homo sapi
29		46	9.3	3025	6	AX598989	AX598989 Sequence
30		46	9.3	3025	6	AX767505	AX767505 Sequence
31		46	9.3	3025	6	AX795863	AX795863 Sequence
32		46	9.3	3025	6	AX822383	AX822383 Sequence
33		46	9.3	3025	6	AX826023	AX826023 Sequence
34	c	46	9.3	151900	2	BX537336	BX537336 Danio rer
35		45.8	9.3	165398	2	AC104677	AC104677 Sus scrof
36		45.6	9.2	6668	6	AX346599	AX346599 Sequence
37		45.6	9.2	7442	6	AX251441	AX251441 Sequence
38		45.6	9.2	66993	2	AC138074	AC138074 Homo sapi
39		45.6	9.2	170627	2	AC125567	AC125567 Rattus no
40		45.6	9.2	236566	2	AC094427	AC094427 Rattus no
41		45.4	9.2	184535	2	CNS05TCJ	AL355100 Homo sapi
42		45.2	9.1	88549	3	AC116979	AC116979 Dictyoste
43	c	45	9.1	127709	2	CNS05TC1	AL352983 Homo sapi
44	c	45	9.1	178948	5	AL929434	AL929434 Zebrafish
45		44.8	9.1	5208	6	AX345825	AX345825 Sequence

ALIGNMENTS

RESULT 1
AR182672
LOCUS AR182672 494 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 7 from patent US 6338961.
ACCESSION AR182672
VERSION AR182672.1 GI:20225879
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 494)
AUTHORS DeRose,R., Chaubet,N. and Gigot,C.
TITLE Isolated DNA sequence capable of serving as regulatory element in a chimeric gene which can be used for the transformation of plants
JOURNAL Patent: US 6338961-A 7 15-JAN-2002;

```
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             100.0%;  Score 494;  DB 6;  Length 494;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-105;
  Matches  494;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480

Qy        481 TCCCGGGATCTGCG 494
              ||||||||||||||
Db        481 TCCCGGGATCTGCG 494


RESULT 2
A59350
LOCUS       A59350                   494 bp    DNA     linear   PAT 06-MAR-1998
DEFINITION  Sequence 7 from Patent WO9704114.
ACCESSION   A59350
VERSION     A59350.1  GI:3714676
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Derose,R., Chaubet,N. and Gigot,C.
  TITLE     ISOLATED DNA SEQUENCE FOR USE AS A REGULATOR REGION IN A CHIMERIC
            GENE USEFUL FOR TRANSFORMING PLANTS
  JOURNAL   Patent: WO 9704114-A 7 06-FEB-1997;
            RHONE POULENC AGROCHIMIE (FR)
COMMENT     Other publication FR 2736929 970124.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Arabidopsis thaliana"
                     /mol_type="unassigned DNA"
                     /strain="COLUMBIA"
                     /db_xref="taxon:3702"
                     /clone="C22"
                     /clone_lib="COSMID"
ORIGIN

  Query Match             99.4%;  Score 490.8;  DB 6;  Length 494;
  Best Local Similarity   99.6%;  Pred. No. 1.7e-104;
  Matches  492;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 180

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        421 TTTAGTTATGGATAATGAAAATCCCCAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480

Qy        481 TCCCGGGATCTGCG 494
              ||||||||||||||
Db        481 TCCCGGGATCTGCG 494


RESULT 3
ATH3G
LOCUS       ATH3G                    4833 bp    DNA     linear   PLN 08-JAN-1996
DEFINITION  A.thaliana H3 gene 1 and H3 gene 2 for H3.3-like histone variant.
ACCESSION   X60429
VERSION     X60429.1  GI:16323
KEYWORDS    histone; histone H3; histone H3.3 homologue.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 4833)
  AUTHORS   Chaubet,N., Clement,B. and Gigot,C.
  TITLE     Genes encoding a histone H3.3-like variant in Arabidopsis contain
            intervening sequences
  JOURNAL   J. Mol. Biol. 225 (2), 569-574 (1992)
  MEDLINE   92277663
   PUBMED   1593639
REFERENCE   2  (bases 1 to 4833)
  AUTHORS   Gigot,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-1991) C. Gigot, Inst de Biologie Mol des Plantes,
            12 Rue du General Zimmer, 67084 Strasbourg Cedex, FRANCE
FEATURES             Location/Qualifiers
     source          1. .4833
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /strain="Columbia"
```

```
                /db_xref="taxon:3702"
                /cell_line="Columbia"
                /clone_lib="cosmid c22"
                /dev_stage="callus"
     gene       691. .1100
                /gene="H3, gene1"
     exon       <691
                /gene="H3, gene1"
                /number=1
     intron     692. .1100
                /gene="H3, gene1"
                /number=1
     gene       join(1101. .1172,1260. .1336,1431. .1521,1620. .1790)
                /gene="H3,gene1"
     CDS        join(1101. .1172,1260. .1336,1431. .1521,1620. .1790)
                /gene="H3,gene1"
                /codon_start=1
                /product="histone H3.3 like protein"
                /protein_id="CAA42957.1"
                /db_xref="GI:16324"
                /db_xref="GOA:P59169"
                /db_xref="SWISS-PROT:P59169"
                /translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
                GTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAYLV
                GLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA"
     gene       1101. .1172
                /gene="H3, gene1"
     exon       1101. .1172
                /gene="H3, gene1"
                /number=2
     gene       1173. .1259
                /gene="H3, gene1"
     intron     1173. .1259
                /gene="H3, gene1"
                /number=2
     gene       1260. .1336
                /gene="H3, gene1"
     exon       1260. .1336
                /gene="H3, gene1"
                /number=3
     gene       1337. .1430
                /gene="H3, gene1"
     intron     1337. .1430
                /gene="H3, gene1"
                /number=3
     gene       1431. .1521
                /gene="H3, gene1"
     exon       1431. .1521
                /gene="H3, gene1"
                /number=4
     gene       1522. .1619
                /gene="H3, gene1"
     intron     1522. .1619
                /gene="H3, gene1"
                /number=4
     gene       1620. .1790
                /gene="H3, gene1"
     exon       1620. .>1790
                /gene="H3, gene1"
                /number=5
     gene       join(2912. .2983,3065. .3141,3227. .3488)
                /gene="H3,gene 2"
     CDS        join(2912. .2983,3065. .3141,3227. .3488)
                /gene="H3,gene 2"
                /codon_start=1
                /product="histone H3.3 like protein"
                /protein_id="CAA42958.1"
                /db_xref="GI:404825"
                /db_xref="GOA:P59169"
                /db_xref="SWISS-PROT:P59169"
                /translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
                GTVALREIRKYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSHAVLALQEAAEAYLV
                GLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA"
     gene       2912. .2983
                /gene="H3, gene 2"
     exon       <2912. .2983
                /gene="H3, gene 2"
                /number=1
     gene       2984. .3064
                /gene="H3, gene 2"
     intron     2984. .3064
                /gene="H3, gene 2"
                /number=1
     gene       3065. .3141
                /gene="H3, gene 2"
     exon       3065. .3141
                /gene="H3, gene 2"
                /number=2
     gene       3142. .3226
                /gene="H3, gene 2"
     intron     3142. .3226
                /gene="H3, gene 2"
                /number=2
     gene       3227. .3488
                /gene="H3, gene 2"
     exon       3227. .>3488
                /gene="H3, gene 2"
                /number=3
ORIGIN

  Query Match            96.7%;  Score 477.8;  DB 8;  Length 4833;
  Best Local Similarity  99.6%;  Pred. No. 1.6e-101;
  Matches  479;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2433 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 2492

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2493 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 2552

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       2553 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 2612

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2613 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 2672

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2673 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 2732

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2733 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 2792

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2793 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 2852

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db       2853 TTTAGTTATGGATAATGAAAATCCCCAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 2912

Qy        481 T 481
              |
Db       2913 T 2913


RESULT 4
ATCHRIV92/c
LOCUS       ATCHRIV92             112067 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 92.
ACCESSION   AL161596
```

```
VERSION     AL161596.2  GI:7271037
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 107700)
  AUTHORS   Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 9546 to 9977)
  AUTHORS   Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 107578 to 112067)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 112067)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV91 at the 5' end and an
            overlap with ATCHRIV93 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .112067
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            6122. .8203
                     /gene="AT4g39830"
     CDS             join(6122. .6415,6481. .6573,6652. .6864,6959. .7153,
                     7250. .8203)
                     /gene="AT4g39830"
                     /note="strong similarity to L-ascorbate oxidase, Cucumis
                     sativus, PIR1:KSKVAO
                     Contains Glycosyl hydrolases family 5 signature
                     AA384-393;Multicopper oxidases signatures
                     AA131-151;Multicopper oxidases signatures
                     AA540-560;Multicopper oxidases signatures
                     AA545-556;Prokaryotic membrane lipoprotein lipid
                     attachment site AA567-577
                     contains EST gb:AI994215.1, AA585817, AI996008.1"
                     /codon_start=1
                     /product="putative L-ascorbate oxidase"
                     /protein_id="CAB80646.1"
                     /db_xref="GI:7271038"
                     /db_xref="GOA:O65670"
                     /db_xref="SPTREMBL:O65670"
                     /translation="MMRPKRSSDTVHVFNLMVLCFIALFFSSVLCQGKIRRFKWEVKY
                     EFKSPDCFEKLVITINGKFPGPTIKAQQGDTIVVELKNSFMTENVAVHWHGIRQIGTP
                     WFDGVEGVTQCPILPGEVFIYQFVVDRPGTYMYHSHYGMQRESGLIGMIQVSPPATEP
                     EPFTYDYDRNFLLTDWYHKSMSEKATGLASIPFKWVGEPQSLMIQGRGRFNCSNNLTT
                     PPSLVSGVCNVSNADCSRFILTVIPGKTYRLRIGSLTALSALSFQIEGHNLTVVEADG
                     HYVEPFTVKNLFVYSGETYSVLLKADQNPRRNYWITSSIVSRPATTPPATAVLNYYPN
                     HPRRRPPTSESSNIVPEWNDTRSRLAQSLAIKARRGFIHALPENSDKVIVLLNTQNEV
                     NGYRRWSVNNVSYHHPKTPYLIALKQNLTNAFDWRFTAPENYDSRNYDIFAKPLNANA
                     TTSDGIYRLRFNSTVDVILQNANTMNANNSETHPWHLHGHDFWVLGYGEGKFNESEDP
                     KRYNRVDPIKKNTVAVQPFGWTALRFRADNPGVWSFHCHIESHFFMGMGIVFESGIDK
                     VSSLPSSIMGCGQTKR"
     exon            6122. .6415
                     /gene="AT4g39830"
                     /number=1
     intron          6416. .6480
                     /gene="AT4g39830"
                     /number=1
     exon            6481. .6573
                     /gene="AT4g39830"
                     /number=2
     intron          6574. .6651
                     /gene="AT4g39830"
                     /number=2
     exon            6652. .6864
                     /gene="AT4g39830"
                     /number=3
     intron          6865. .6958
                     /gene="AT4g39830"
                     /number=3
     exon            6959. .7153
                     /gene="AT4g39830"
                     /number=4
     intron          7154. .7249
                     /gene="AT4g39830"
                     /number=4
     exon            7250. .8203
                     /gene="AT4g39830"
                     /number=5
     gene            12287. .13642
                     /gene="AT4g39840"
     CDS             12287. .13642
                     /gene="AT4g39840"
                     /note="contains EST gb:AI999725.1, Z46537, F13826,
                     AI996686.1"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80647.1"
                     /db_xref="GI:7271039"
                     /db_xref="SPTREMBL:Q9SMR9"
                     /translation="MSPLKYLCFRKVLLLLLYSSLFLFSSSSFFASSVDDHELQRIDP
                     HFRVRRLLVKDLDSTGDDDETNLPPPPKKKKLLTDSVSSSSSSSGTKKNQTKLLKPIS
                     SSSSTKNQTKLAKTTTMGTSHKLNSTKSSSNTTKTSSELKKLNSGTKSTNSTSSIKKS
                     ADLSKSSSSKNKTTIKPPSSKLSSPPSEKKSQPSSKPVTKSKQSEKEIKPFWLDDEED
                     EDFVSEFRDLPTRPQRSLIPDLERISTTSKNYINKANKQITKNFKPYFGNKYAPTIAS
                     VVSFVFILVPLLLVSLIFNRFKAYFSLQKILIFIQIYLSIYFSILCLSSLVTGIEPLK
                     FLYATSSSTYVCLQILQTLGYVFYLLLLLMYLVLVFSTDCGLGLKVLGLAQTFVGFAV
                     GLHYYVAVFHRVVLRQPPKTNWKIHGVYATCFLLICLLSNAERRKKEYLEEGGDEGKK
                     N"
     exon            12287. .13642
                     /gene="AT4g39840"
                     /number=1
     misc_feature    15539. .16205
                     /note="EST N95877 matches to positions 26106-26355,
                     26723-26772; possibly related to subsequent gen"
     gene            16239. .23312
                     /gene="AT4g39850"
     CDS             join(16239. .16300,16394. .16679,16981. .17205,17341. .17538,
                     17675. .18129,18222. .18357,18433. .18511,19039. .19271,
                     19326. .19630,19712. .19844,19901. .20056,20213. .20435,
                     20515. .20623,20845. .20935,21205. .21372,21453. .21568,
                     21626. .21860,21924. .22104,22162. .22362,22534. .22680,
                     22750. .22983,23134. .23312)
                     /gene="AT4g39850"
                     /note="similarity to 70K peroxisomal membrane protein,
                     Rattus norvegicus , PIR2:A35723
                     Contains ABC transporters family signature AA580-594;ABC
                     transporters family signature AA1369-1383;ATP/GTP-binding
                     site motif A (P-loop) AA1247-1254
                     contains EST gb:N95877"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB80648.1"
                     /db_xref="GI:7271040"
                     /db_xref="GOA:Q9SMR8"
                     /db_xref="SPTREMBL:Q9SMR8"
                     /translation="MPSLQLLQLTERGRGLVASRRKSILLAAGIVAAGGTAVYLKSRV
```

```
                     ASRRPDSSRLCNGQSDDDETLEKLTATDQNAKITTKKKKGGGLKSLQVLTAILLSQMG
                     KMGARDLLALVATVVFRTALSNRLAKVQGFLFRAAFLRRAPLFLRLISENIMLCFMLS
                     TLHSTSKYITGALSLRFRKILTKIIHSHYFENMVYYKISHVDGRITHPEQRIASDVPR
                     FSSELSDLILDDLTAVTDGILYAWRLCSYASPKYIFWILAYVLGAGTAIRNFSPSFGK
                     LMSKEQQLEGEYRQLHSRLRTHSESIAFYGGETREESHIQQKFKNLVSHMSHVLHDHW
                     WFGMIQDFLLKYLGATVAVILIIEPFFSGHLRPDDSTLGRAEMLSNIRYHTSVIISLF
                     QALGTLSISSRRLNRLSGYADRIHELMAVSRELSGDDKSSFQRNRSRNYLSEANYVEF
                     SDVKVVTPTGNVLVEDLTLRVEQGSNLLITGGLWPLVSGHIVKPGVGSDLNKEIFYVP
                     QRPYMAVGTLRDQLIYPLTSGQESELLTEIGMVELLKNVSILVSLFYTRMVDLEYLLD
                     RYQPEKEVNWGDELSLGEQQRLGMARLFYHKPKFAILDECTSAVTTDMEERFAAKVRA
                     MGTSCITISHRPALVAFHDVVLSLDGEGGWSVHYKRDDSALLTDAEIDSVKSSDTDRQ
                     NDAMVVQRAFAAARKVSAEVLFESATNSKAQSYQTQLIARSPVVDKSVVLPRFPQPQT
                     SQRALPSRVAAMLNVLLPWKPHLLFCISTWLTYSSCADANIDIILLKITDHFIVTFQI
                     PTIFDKQGAQLLAVACLVVSRTLISDRIASLNGTTVKYVLEQDKAAFVRLIGLSVLQS
                     GASSIIAPSLRHLTQRLALGWRIRLTQHLLRNYLRNNAFYKVFHMSGNSIDADQRLTR
                     DLEKLTADLSGLLTGMVKPSVDILWYVRYVGLISNICFMRGVAILYTYMLLGLGFLRR
                     VAPDFGDLAGEEQQLEGKFRPTSSCLCRFMHERLNTHAESIAFFGGGAREKAVSFLIA
                     LAIAAGFWVVIFESKLPEVNLELFYHLCDDGLFMSNVASLMVDKKFRALLDHSLMLLR
                     KKWLYGILDDFVTKQLPNNVTWGLSLLYALEHKGDRALVSTQVKYLHVLYMMLLIPGE
                     LAHALRYLASVVSQSFMAFGDILELHKKFLELSGGINRIFELDEFLDASQSGVTSENQ
                     TSRLDSQDLLSFSEVDIITPAQKLMASKLSCEIVSGKSLLVTGPNGSGKTSVFRVLRD
                     IWPTVCGRLTKPSLDIKELGSGNGMFFVPQRPYTCLGTLRDQIIYPLSKEEAEKRAAK
                     LYTSGESSTEAGSILDSHLKTILENVRLVYLLERDVGGWDATTNWEDILSLGEQQRLG
                     MVSLL"
     exon            16239. .16300
                     /gene="AT4g39850"
                     /number=1
     intron          16301. .16393
                     /gene="AT4g39850"
                     /number=1
     exon            16394. .16679
                     /gene="AT4g39850"
                     /number=2
     intron          16680. .16980
                     /gene="AT4g39850"
                     /number=2
     exon            16981. .17205
                     /gene="AT4g39850"
                     /number=3
     intron          17206. .17340
                     /gene="AT4g39850"
                     /number=3
     exon            17341. .17538
                     /gene="AT4g39850"
                     /number=4
     intron          17539. .17674
                     /gene="AT4g39850"
                     /number=4
     exon            17675. .18129
                     /gene="AT4g39850"
                     /number=5

  Query Match             96.7%;  Score 477.8;  DB 8;  Length 112067;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-101;
  Matches  479;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83915 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 83856

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83855 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 83796

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      83795 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 83736

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83735 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 83676

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83675 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 83616

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83615 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 83556

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83555 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 83496

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      83495 TTTAGTTATGGATAATGAAAATCCCCAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 83436

Qy        481 T 481
              |
Db      83435 T 83435
```

```
RESULT 5
ATT5J17/c
LOCUS       ATT5J17               118267 bp    DNA     linear   PLN 21-SEP-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone T5J17 (ESSA
            project).
ACCESSION   AL035708
VERSION     AL035708.2  GI:5918309
KEYWORDS    .
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Bevan,M., Murphy,G., Ridley,P., Hudson,S., Bancroft,I., Mewes,H.W.,
            Mayer,K.F.X., Lemcke,K. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118267)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     On Sep 22, 1999 this sequence version replaced gi:4490734.
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .118267
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     misc_feature    1. .22789
                     /note="position 1-22789 overlaps to BAC T19P19, EMBL
                     accession AL022605; please refer to this entry for
                     analysis and annotation"
     gene            22854. .24209
                     /gene="T5J17.10"
     CDS             22854. .24209
                     /gene="T5J17.10"
                     /note="contains EST gb:F13826, Z46537"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB38897.1"
                     /db_xref="GI:4490735"
                     /db_xref="SPTREMBL:Q9SMR9"
                     /translation="MSPLKYLCFRKVLLLLLYSSLFLFSSSSFFASSVDDHELQRIDP
                     HFRVRRLLVKDLDSTGDDDETNLPPPPKKKKLLTDSVSSSSSSSGTKKNQTKLLKPIS
```

```
                SSSSTKNQTKLAKTTTMGTSHKLNSTKSSSNTTKTSSELKKLNSGTKSTNSTSSIKKS
                ADLSKSSSSKNKTTIKPPSSKLSSPPSEKKSQPSSKPVTKSKQSEKEIKPFWLDDEED
                EDFVSEFRDLPTRFQRSLIPDLERISTTSKNYINKANKQITKNFKPYFGNKYAPTIAS
                VVSFVFILVPLLLVSLIFNRFKAYFSLQKILIFIQIYLSIYFSILCLSSLVTGIEPLK
                FLYATSSSTYVCLQILQTLGYVFYLLLLLMYLVLVFSTDCGLGLKVLGLAQTFVGFAV
                GLHYYVAVFHRVVLRQPPKTNWKIHGVYATCFLLICLLSNAERRKKEYLEEGGDEGKK
                N"
exon            22854. .24209
                /gene="T5J17.10"
                /number=1
misc_feature    26106. .26772
                /note="EST N95877 matches to positions 26106-26355,
                26723-26772; possibly related to subsequent gen"
gene            26806. .33879
                /gene="T5J17.20"
CDS             join(26806. .26867,26961. .27246,27548. .27772,27908. .28105,
                28242. .28696,28789. .28924,29000. .29078,29606. .29838,
                29893. .30197,30279. .30411,30468. .30623,30780. .31002,
                31082. .31190,31412. .31502,31772. .31939,32020. .32135,
                32193. .32427,32491. .32671,32729. .32929,33101. .33247,
                33317. .33550,33701. .33879)
                /gene="T5J17.20"
                /note="similarity to 70K peroxisomal membrane protein,
                Rattus norvegicus , PIR2:A35723
                Contains ABC transporters family signature
                [LSLGEQQRLGMARLF][LSLGEQQRLGMVSLL]"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB38898.1"
                /db_xref="GI:4490736"
                /db_xref="GOA:Q9SMR8"
                /db_xref="SPTREMBL:Q9SMR8"
                /translation="MPSLQLLQLTERGRGLVASRRKSILLAAGIVAAGGTAVYLKSRV
                ASRRPDSSRLCNGQSDDDETLEKLTATDQNAKITTKKKKGGGLKSLQVLTAILLSQMG
                KMGARDLLALVATVVFRTALSNRLAKVQGFLFRAAFLRRAPLFLRLISENIMLCFMLS
                TLHSTSKYITGALSLRFRKILTKIIHSHYFENMVYYKISHVDGRITHPEQRIASDVPR
                FSSELSDLILDDLTAVTDGILYAWRLCSYASPKYIFWILAYVLGAGTAIRNFSPSFGK
                LMSKEQQLEGEYRQLHSRLRTHSESIAFYGGETREESHIQQKFKNLVSHMSHVLHDHW
                WFGMIQDFLLKYLGATVAVILIIEPFFSGHLRPDDSTLGRAEMLSNIRYHTSVIISLF
                QALGTLSISSRRLNRLSGYADRIHELMAVSRELSGDDKSSFQRNRSRNYLSEANYVEF
                SDVKVVTPTGNVLVEDLTLRVEQGSNLLITGGLWPLVSGHIVKPGVGSDLNKEIPYVP
                QRPYMAVGTLRDQLIYPLTSGQESELLTEIGMVELLKNVSILVSLFYTRMVDLEYLLD
                RYQPEKEVNWGDELSLGEQQRLGMARLFYHKPKFAILDECTSAVTTDMEERFAAKVRA
                MGTSCITISHRPALVAFHDVVLSLDGEGGWSVHYKRDDSALLTDAEIDSVKSSDTDRQ
                NDAMVVQRAFAAARKVSAEVLFESATNSKAQSYQTQLIARSPVVDKSVVLPRFPQPQT
                SQRALPSRVAAMLNVLLPWKPHLLFCISTWLTYSSCADANIDIILLKITDHFIVTFQI
                PTIFDKQGAQLLAVACLVVSRTLISDRIASLNGTTVKYVLEQDKAAFVRLIGLSVLQS
                GASSIIAPSLRHLTQRLALGWRIRLTQHLLRNYLRNNAFYKVFHMSGNSIDADQRLTR
                DLEKLTADLSGLLTGMVKPSVDILWYVRYVGLISNICFMRGVAILYTYMLLGLGFLRR
                VAPDFGDLAGEEQQLEGKFRPTSSCLCRFMHERLNTHAESIAFFGGGAREKAVSFLIA
                LAIAAGFWVVIFESKLPEVNLELFYHLCDDGLFMSNVASLMVDKKFRALLDHSLMLLR
                KKWLYGILDDFVTKQLPNNVTWGLSLLYALEHKGDRALVSTQVKYLHVLYMMLLIPGE
                LAHALRYLASVVSQSFMAFGDILELHKKFLELSGGINRIFELDEFLDASQSGVTSENQ
                TSRLDSQDLLSFSEVDIITPAQKLMASKLSCEIVSGKSLLVTGPNGSGKTSVFRVLRD
                IWPTVCGRLTKPSLDIKELGSGNGMFFVPQRPYTCLGTLRDQIIYPLSKEEAEKRAAK
                LYTSGESSTEAGSILDSHLKTILENVRLVYLLERDVGGWDATTNWEDILSLGEQQRLG
                MVSLL"
exon            26806. .26867
                /gene="T5J17.20"
                /number=1
intron          26868. .26960
                /gene="T5J17.20"
                /number=1
exon            26961. .27246
                /gene="T5J17.20"
                /number=2
intron          27247. .27547
                /gene="T5J17.20"
                /number=2
exon            27548. .27772
                /gene="T5J17.20"
                /number=3
intron          27773. .27907
                /gene="T5J17.20"
                /number=3
exon            27908. .28105
                /gene="T5J17.20"
                /number=4
intron          28106. .28241
                /gene="T5J17.20"
                /number=4
exon            28242. .28696
                /gene="T5J17.20"
                /number=5
intron          28697. .28788
                /gene="T5J17.20"
                /number=5
exon            28789. .28924
                /gene="T5J17.20"
                /number=6
intron          28925. .28999
                /gene="T5J17.20"
                /number=6
exon            29000. .29078
                /gene="T5J17.20"
                /number=7
intron          29079. .29605
                /gene="T5J17.20"
                /number=7
exon            29606. .29838
                /gene="T5J17.20"
                /number=8
intron          29839. .29892
                /gene="T5J17.20"
                /number=8
exon            29893. .30197
                /gene="T5J17.20"
                /number=9
intron          30198. .30278
                /gene="T5J17.20"
                /number=9
exon            30279. .30411
                /gene="T5J17.20"
                /number=10
intron          30412. .30467
                /gene="T5J17.20"
                /number=10
exon            30468. .30623
                /gene="T5J17.20"
                /number=11
intron          30624. .30779
                /gene="T5J17.20"
                /number=11
exon            30780. .31002
                /gene="T5J17.20"
                /number=12
intron          31003. .31081
                /gene="T5J17.20"
                /number=12
exon            31082. .31190
                /gene="T5J17.20"
                /number=13
intron          31191. .31411
                /gene="T5J17.20"
                /number=13
exon            31412. .31502
                /gene="T5J17.20"
                /number=14
intron          31503. .31771
                /gene="T5J17.20"
                /number=14
exon            31772. .31939
                /gene="T5J17.20"
                /number=15
intron          31940. .32019
                /gene="T5J17.20"
```

```
                     /number=15
     exon            32020. .32135
                     /gene="T5J17.20"

  Query Match             96.7%;  Score 477.8;  DB 8;  Length 118267;
  Best Local Similarity   99.6%;  Pred. No. 1.2e-101;
  Matches  479;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94482 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 94423

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94422 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 94363

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      94362 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 94303

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94302 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 94243

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94242 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 94183

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94182 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 94123

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      94122 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 94063

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db      94062 TTTAGTTATGGATAATGAAAATCCCCAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 94003

Qy        481 T 481
              |
Db      94002 T 94002
```

```
RESULT 6
AJ592632
LOCUS       AJ592632               777 bp    DNA     linear   PLN 23-OCT-2003
DEFINITION  Arabidopsis thaliana T-DNA flanking sequence, right border, clone
            627A08.
ACCESSION   AJ592632
VERSION     AJ592632.1  GI:37942256
KEYWORDS    right border; T-DNA flanking sequence.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1
  AUTHORS   Brunaud,V., Balzergue,S., Dubreucq,B., Aubourg,S., Samson,F.,
            Chauvin,S., Bechtold,N., Cruaud,C., DeRose,R., Pelletier,G.,
            Lepiniec,L., Caboche,M. and Lecharny,A.
  TITLE     T-DNA integration into the Arabidopsis genome depends on sequences
            of pre-insertion sites
  JOURNAL   EMBO Rep. 3 (12), 1152-1157 (2002)
  MEDLINE   22363535
   PUBMED   12446565
REFERENCE   2  (bases 1 to 777)
  AUTHORS   Balzergue,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue
            Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT     PCR was performed on DNA from transformants of Arabidopsis thaliana
            plants from INRA (Versailles). The DNA fragment(s) resulting from
            the PCR were directly sequenced from the left or the right border
            to determine the genomic sequence flanking the insertion. T-DNA
            derived sequences were removed. Information to order the
            corresponding mutant line and a link to a database providing a
            graphical display of the insertion site are available at
            http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
            been generated in the framework of the French plant genomics
            program 'Genoplante' (http://www.genoplante.com and
            http://genoplante-info.infobiogen.fr).
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /cultivar="Wassillewskija"
                     /db_xref="taxon:3702"
                     /clone="627A08"
                     /clone_lib="Arabidopsis thaliana T-DNA insertion lines"
     misc_feature    1. .777
                     /note="T-DNA flanking sequence
                     right border"
ORIGIN
```

```
  Query Match             96.4%;  Score 476.2;  DB 8;  Length 777;
  Best Local Similarity   99.4%;  Pred. No. 4.3e-101;
  Matches  478;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 CTCAGGCGAAGAACAGGTATGATTTGTTTGTAATTAGATCAAGGGTTTAGGTCTTTCCAT 196

Qy         61 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 TACTTTTTAATGTTTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTT 256

Qy        121 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGA 180
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        257 CGGGTTTTGTCCCATTCCAGTTTGAAAATAAAGGTCCGTCTTTTAAGTTTGCTGGATCGA 316

Qy        181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTT 376

Qy        241 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 TCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCG 436

Qy        301 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 TGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGAT 496

Qy        361 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 AACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAA 556

Qy        421 TTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 480
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        557 TTTAGTTATGGATAATGAAAATCCCCAATTGACTGTTCAATTTCTTGTTAAATGCGCAGA 616

Qy        481 T 481
              |
Db        617 T 617
```

```
RESULT 7
AX348370
LOCUS       AX348370               12870 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 65 from Patent WO0202806.
ACCESSION   AX348370
VERSION     AX348370.1  GI:18614406
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
```

```
         artificial sequences.
REFERENCE    1
  AUTHORS    Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE      Method and nucleic acids for pharmacogenomic methylation analysis
  JOURNAL    Patent: WO 0202806-A 65 10-JAN-2002;
             Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1..12870
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             11.3%;  Score 55.8;  DB 6;  Length 12870;
  Best Local Similarity   46.5%;  Pred. No. 0.0061;
  Matches  180;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db       3147 TTTATTTGGTTTTAAATGTTTTTATATTTAAGTTTTTTAATTGATTGATAGTTTTTTT 3206
Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db       3207 TATAGTTTTTTAATATAGTTTTGTTTTTTTATTGGTTTTTAGATATGTTTTGTTCGTTTT 3266
Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db       3267 TATTTTGTGTTTTTATTTATATTGTTTTTTTTTTTTGGAATGTTAATTCGGTTTTTTTTT 3326
Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db       3327 TTTTAGTTTTTATTTGTTTTTTATTTTTTTATTTAAATTTTATTTTTTTTAGGAAGTTTT 3386
Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db       3387 TTTTGATTATTTTAGTTTATTTTTATTATTTTTTTTTTTGGATTTTTTTAGGAATTAAAG 3446
Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db       3447 TTAGTGAATACGTATTGATATTTTTTATAATAATTTATTGGGTTTTGGTTTTTTTTTTTT 3506
Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTAT 404
Db       3507 TAGAAGTTCGGTAGGGTATGATAGTTT 3533

RESULT 8
AX348661
LOCUS       AX348661             12870 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 119 from Patent WO0202807.
ACCESSION   AX348661
VERSION     AX348661.1  GI:18614696
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with cell signalling
  JOURNAL   Patent: WO 0202807-A 119 10-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1..12870
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             11.3%;  Score 55.8;  DB 6;  Length 12870;
  Best Local Similarity   46.5%;  Pred. No. 0.0061;
  Matches  180;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db       3147 TTTATTTGGTTTTAAATGTTTTTATATTTAAGTTTTTTAATTGATTGATAGTTTTTTT 3206
Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db       3207 TATAGTTTTTTAATATAGTTTTGTTTTTTTATTGGTTTTTAGATATGTTTTGTTCGTTTT 3266
Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db       3267 TATTTTGTGTTTTTATTTATATTGTTTTTTTTTTTTGGAATGTTAATTCGGTTTTTTTTT 3326
Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db       3327 TTTTAGTTTTTATTTGTTTTTTATTTTTTTATTTAAATTTTATTTTTTTTAGGAAGTTTT 3386
Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db       3387 TTTTGATTATTTTAGTTTATTTTTATTATTTTTTTTTTTGGATTTTTTTAGGAATTAAAG 3446
Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db       3447 TTAGTGAATACGTATTGATATTTTTTATAATAATTTATTGGGTTTTGGTTTTTTTTTTTT 3506
Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTAT 404
Db       3507 TAGAAGTTCGGTAGGGTATGATAGTTT 3533

RESULT 9
SKORFS/c
LOCUS       SKORFS               14281 bp    DNA     linear   PLN 06-NOV-1996
DEFINITION  S.kluyveri linear plasmid pSKL DNA for open reading frames 1-10.
ACCESSION   X54850
VERSION     X54850.1  GI:4868
KEYWORDS    linear plasmid; plasmid.
SOURCE      Saccharomyces kluyveri
  ORGANISM  Saccharomyces kluyveri
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1
  AUTHORS   Hishinuma,F. and Hirai,K.
  TITLE     Genome organization of the linear plasmid, pSKL, isolated from
            Saccharomyces kluyveri
  JOURNAL   Mol. Gen. Genet. 226 (1-2), 97-106 (1991)
  MEDLINE   91238725
   PUBMED   2034232
REFERENCE   2  (bases 1 to 14281)
  AUTHORS   Hishinuma,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1990) Hishinuma F., Mitsubishi Kasei Institute of
            Life Sciences, 11 Minamiooya Machida-shi, Tokyo 194, Japan
FEATURES             Location/Qualifiers
     source          1..14281
                     /organism="Saccharomyces kluyveri"
                     /mol_type="genomic DNA"
                     /strain="UVC40"
                     /db_xref="taxon:4934"
                     /plasmid="pSKL"
     repeat_unit     1..483
                     /note="terminal"
                     /rpt_type=INVERTED
     mRNA            558..>1279
                     /note="ORF 1"
     CDS             578..1279
                     /note="unnamed protein product; ORF 1"
                     /codon_start=1
                     /protein_id="CAA38620.1"
                     /db_xref="GI:4869"
                     /db_xref="GOA:Q04318"
                     /db_xref="SPTREMBL:Q04318"
                     /translation="MEMDDRLFSAYSGDYKTYYSGYISGEMSELEVACIKNGCFLNIK
```

FSFETLKRFSHSRPDLCGYAVSKFISKNYINDYLIEKYKPFYVWYPSFCSYESYLYLL
ENFKFLKYNVGKAIIMLHYNDLYFSYDFEPEMELFVSASVFNNKEVYEDQLSKAKEIG
YFYKYLDIDSESSLEVPLKENIETDFGFDPNTVYGENELCEEYERKFFRSFEFNKGIE
GVKELDEVVRRMGGI"

mRNA 581. .>1279
/note="ORF 1"

mRNA complement(<1276. .4278)
/note="ORF 2"

CDS complement(1276. .4275)
/note="unnamed protein product; ORF 2, has similarity to DNA polymerase"
/codon_start=1
/protein_id="CAA38621.1"
/db_xref="GI:4870"
/db_xref="GOA:Q09038"
/db_xref="SPTREMBL:Q09038"
/translation="MNKELEFLKNEVSYTETEDLDFLKSEVFKIHPYTRSIEKLEKNS
TENSIRSTIEYNQYGQEIYSKLSSNQIELEEAKQLMHKKFIQLRNKNRVPLKRKRKKK
LKFDHDVETLIIHKKYKIDLNLCNMDETQTFTHKDISIIIPTRKSILKHIHVDRKYNY
YCFVTLTLTGPTIFGGVTSKVPILNVQGIPNNDYDERMTKALNDIIERIPDYEEKESN
IVCTRILITITQATWKSIDNWYPKTTRKFKKFKDFTLFLASNRNENCLKQCVEKLGGK
WNYNKKLEDMLPQKKIITYIPALMDIQYVNSMEDLITDYNEEIYSLDCKNIARLLKWN
GHIGVITKLDPTRKIGKIQKQKRLEVEDTNTEEIFFDIESFTNEKNFQIPYLICWSKT
NGSIHKRIGKNCINDFVQYIINLKKNVILYAWFGSGYDYQHVLPYFKKKCIKDKYIIK
NNMITYSELYFKHSKIILKDPFLFILTSLDKAAKAFNVINKGSFPHSLIKDWPDLQKI
YPHWIKTQRKIIEFKHDNKLNIKINTTLEEDNITNNSTIIEKAIEYCSIDVIAMKQVW
QKFKKLVKENLQISISEKTFTLSQLSMKIMEASLDRNIELYVPTSEEYQFIRNSIYGG
RVAAKNGTYNEEIVYADVVSLYPSAMKLLEHSYGKPSKVFSINFNKHGIYDVTLIHKS
NSKPNHYLEFVPRRIDKKLIWSWFKEHRGTYHTYDLLIAIQEGFEIICHSGIEYPNKG
YIFNNFIDKLYTLKDIHTNCKCKEQPCPIRMVAKIALNGGGYGKFVQKPIDKEVYIVK
RDVVAGECENLSQNENGEICIGKRIVKKPKFFNLDGEDYDKMIIERDDDPIYSTQCGV
SILSASRYRLYNLCKKFEGLEIIYSDTDSIFVRKNTIDWELFKNTCGKNLGELDSTID
KTKNAIIHYMLIGGPKMYAFQYKNNKNNIITKLHCKGVPTFMLSIDQFEYLLHKKDRK
LAYHFEIIRRKLVTVTTDKIIKDIKQT"

mRNA complement(<4283. .5995)
/note="ORF 3"

CDS complement(4283. .5977)
/note="unnamed protein product; ORF 3"
/codon_start=1
/protein_id="CAA38622.1"
/db_xref="GI:4871"
/db_xref="GOA:Q09048"
/db_xref="SPTREMBL:Q09048"
/translation="MFKKLAFLANQLESPLVNIELEARMKIQQPIDIIFDQEITITYY
RSTMYPSIVFRRLNNSELNSKELVEKIKFNDFIIALSIEQSYTNSEIKFPLIESNTRK
IYRKRLQEDPIVEITKYNNEFNLEVEFDINNYEKVEKILKEWKSPYWPPVKPMEISSS
NLAKKLSHNDQWCISPKADGEHIIQYINDESSIIIHDNGFISDIDGNEKNNIIPRYVY
EAELMNNKKILYFDCLMYDYKNITKLDYIKRREYLKEKKEIILFNNVYALKSYLYKPA
QYKTDGYIISNIKNKKLVYKSKFINTVDLRYVNGYLLLENEKFSDRIPKDTTYNFEED
KIYEFDIEMNLIKERKDKTIANYRFPYDDNPIYKIINGIGIPTLRYHHNKIKTELLNI
LPKSTLLDIGSGKGGDIHKWYHLKFKKIYAVDPILDLRQHSKNVVEIRDYVQNIYQLI
DYNSVSILFVPWDDIYLDVIQKAKHVVIAWMSNPKNYECDSFKCKVTDKVELNIPHTE
TSNFVIENKTNYKKILKNLKDNGWYHQEIKIKMEIGTIQERELSRMYTYHYLKKIK"

mRNA 6166. .>7927
/note="ORF 4"

CDS 6170. .7927
/note="unnamed protein product; ORF 4"
/codon_start=1
/protein_id="CAA38623.1"
/db_xref="GI:4872"
/db_xref="GOA:Q04332"
/db_xref="SPTREMBL:Q04332"
/translation="MNLFTVSDIKIEQYGFVNTFKNQNFIISYLHPLNPYRSLVICYE
VGLGKTYASACLAHVYLSYGYKVLYLSSSLNSISNFRNEYNKVITDSRLTSYEKNIDV
KSFSKFYNSEKKNENNEEYGLVIVDEVHNLRELAIRYKTIKNKLDSMINSKILIVTAT
PMIDSHKELDSIINLTNEKTKIIFSENKISKDVKINYVGSNINGETLFLSRMKGKQLE
DYYKSLKNKNDTVYTETRQASISWNDEFNPEIPLDEQSSKISRVIETLKDGELTVIFC
FYVKRGIDFMAKVLEHFGYRKWDPLYNGKSYAVIDGRTSHKDVKEILKVFNSIANTNG
KQIQVLIGSSVLSESITLYRIRHLHILSPFWNYGQVKQSIGRGVRIGSHDGLLDKTIN
VYLHAAYAEENFKGKDIDIWETAYRKKVEIDKKLDKLKIENKVENIVEKISVPEVDNN
LVIKINEWVWDLRNCFEQNKFKISWCRVYKDKAIGYNIENNTKIIGNIPDYIKINTPL
KNGYTVWRSCIDNKLRLSYIDEKVNKFTKRGKLLANVNINKIAKDLNCGESTSDIINT
LKKENRYFDKQIEYDLQQY"

mRNA 7832. .>8387
/note="ORF 5"

mRNA 7873. .>8387
/note="ORF 5"

CDS 7908. .8387
/note="unnamed protein product; ORF 5"
/codon_start=1
/protein_id="CAA38624.1"
/db_xref="GI:4873"
/db_xref="GOA:Q04306"
/db_xref="SPTREMBL:Q04306"
/translation="MIYNNIDTDLISLGKAKKFTEEEEKWITPIYYEGESLDFTLKNK
YVKINKIEENSYGKKFVTIKSKEYSKIIESISEKLGTVSPISSDGSFRAPINNKTKIN
EKNLDDVSFNACVSLVFPTIYKDTEKKTLQIYIKDLVVVEIIKDELEIELDKLSLAM"

mRNA 8401. .>11365
/note="ORF 6"

CDS 8408. .11365
/note="unnamed protein product; ORF 6, has similarity to RNA polymerase"
/codon_start=1
/protein_id="CAA38625.1"
/db_xref="GI:4874"
/db_xref="GOA:Q04387"
/db_xref="SPTREMBL:Q04387"
/translation="MEYTDYGQIKIYETLFEDLDVSTIKNTPIECIEMKLSYCGIYNN
RELPIMVGSKFDLHDNILGIKGYFIIDGICKSVNNIKMTERISFSKDRAYLTDNSKIE
IKDMFNYVIKKDNKSRKWSLPINWRDIIKYSSNKKELEDHLSLIYTKSDKVIKNEIDL
MIICYMFECWLGIKKESVHPWRLITPGELIYDIINKGEDVVKCFRTNTWKVKYINNVN
SVSEDMKHYNIVGDIESIRRITIPTQRENTNMENRMVKLEDKYKICPVQTSDGALCGT
VYYLCKDSKVTNTPQKVIEKEEGNYHTFLNSVYIGKTSIEYINKHKDNGDIHKINDIY
YIFNRYGRVIKSNSIVSYTASLIPYRKNNPPIRSMFSCSMMKQAISGENIYSKNLLNT
KSLIFGGLGHDIVVAIMPWYGYNIEDAIVINKKTASKFRSRKCRIYRENNCKIIGIYV
KKNQYIIKGQLLYKTYNPELVKTIEFIYSDQDGTIDELKHSEKYIKLVIIKERDLEVG
DKMTSRHGQKGVVSLILNNNEMPYYYEKNMKKEIDLLINPHAFPSRMTMGQIKEMGNN
EKEVYVNNKKLDNKILVGRCFYMALRHQVDDKIQYRNSGNIDIINKQPVSGKKNNGGL
RFGQMERDILIGLGAWNTLKELWSIDKTNIYVCPKSGKINPKCCKIKRESHQYFKICL
SYMRALGHDILIKNNKYSIIEFDPSDLKKTTTLKFGDLDPLDIRIYRKIVVLPLCLRS
TYLNKLYTKHLDNTKEIEKEIKKLLKSKKGAYHKLVEGHRVDHCIRSVIVPDPTLDIY
TVKIPLGANISTSYGILNRQPSLNVDSMKIVKLIQGKNKTISFNPLLCKSFNADFDGD
EMNIYGIKNEESIQELKQKLIVKEDKTQDYILGKEIINKGLTANRQGIKLMIEKGSKG
KEFNYTHMFEKIGQVTINENIKTYIDQPYNGEIKDDDWYILSMAARQSAASIGVNTPI
TGYLESVCNQMYI"

mRNA complement(<11371. .11777)
/note="ORF 7"

CDS complement(11371. .11769)
/note="unnamed protein product; ORF 7"
/codon_start=1
/protein_id="CAA38626.1"
/db_xref="GI:4875"
/db_xref="GOA:Q09059"
/db_xref="SPTREMBL:Q09059"
/translation="MNEDILNNLNYFSEVTQRTLSAFHTFGSDASLIDEINSIKRFIT
NAKVNKLVKVKRREDGKFIRYFSDNSCDIVDLEDLSKPAYYENKYDFSVIEKALSQNI
NPISSLKLSNSIEHIKNKKIEKPSEIIFFM"

mRNA complement(<11371. .11761)
/note="ORF 7"

mRNA complement(<11766. .12027)
/note="ORF 8"

mRNA complement(<11766. .12011)
/note="ORF 8"

CDS complement(11766. .12011)
/note="unnamed protein product; ORF 8"
/codon_start=1
/protein_id="CAA38627.1"
/db_xref="GI:4876"
/db_xref="GOA:Q04337"
/db_xref="SPTREMBL:Q04337"
/translation="MYAITNVDFLEYYSNDIPKEILSSYLIGSKCNSKGKIKLTLKPG
NEFIQEGPYILPPIRWLDSFEYYAEFVIVCDVDYVEK"

mRNA 11989. .>13389
/note="ORF 9"

CDS 12013. .13389
/note="unnamed protein product; ORF 9"

```
                     /codon_start=1
                     /protein_id="CAA38628.1"
                     /db_xref="GI:4877"
                     /db_xref="GOA:Q04331"
                     /db_xref="SPTREMBL:Q04331"
                     /translation="MEIIEYKIDFIYDKINEYVWSDNFQLSEEWEILRLKKTRTLQKS
                     YMDRSPYTEPNTQYTIALYSQNKWYCLHNGKGNVISPLRKNPEIPNLKIIKSTFTPFL
                     TIYKIQFDFHVMMTLLNLNDEIKVLEFKRPSLTEGIIIYNKMKINVYNSLFSYKGENQ
                     KEFKLFCINTKKMYEKDSPKIIKLYSKISFSDDTHKSIDFRTWYKAKRENQQEWPKIS
                     KIQLENSIEFPKNSNIWYYYPQKFVGLARRNDDNPLQLIPKIYKKDHSKNKNTNLYHY
                     LQTGKPKNDDTIKIPHTLRKGDLKYAKQIKKLTNEKYIEMSYNGEIIETDKNPEYYMY
                     RNILIKKTDIKINIDLDAQLLNQNNERIMILKNNKWIKSQGPRIKNIITIPWNYKQIL
                     HDYNKLGRLKKGKIIDLSNIGIVDENATDIITWPETENLYSSKTLETQRLITFKFQKN

  Query Match             10.2%;  Score 50.6;  DB 8;  Length 14281;
  Best Local Similarity   48.8%;  Pred. No. 0.099;
  Matches  137;  Conservative    0;  Mismatches  144;  Indels    0;  Gaps    0;

Qy        195 TTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGT 254
Db      11290 TTGTCTTGCGGCCATAGATAGTATATACCAATCATCATCTTTTATTTCGCCGTTGTAAGG 11231

Qy        255 TCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTA 314
Db      11230 TTGATCTATATAAGTTTTTATGTTTTCGTTTATAGTTACTTGACCTATTTTTTCGAACAT 11171

Qy        315 GGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGG 374
Db      11170 ATGTGTATAATTAAATTCTTTGCCTTTAGATCCTTTTTCGATCATTAATTTTATGCCTTG 11111

Qy        375 ATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATA 434
Db      11110 TCTATTTGCGGTCAATCCTTTATTTATTATTTCTTTTCCTAATATATAGTCTTGAGTTTT 11051

Qy        435 ATGAAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATGC 475
Db      11050 ATCTTCTTTAACTATTAATTTTTGTTTTAATTCTTGTATAC 11010


RESULT 10
I66494
LOCUS       I66494                   7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             10.0%;  Score 49.6;  DB 6;  Length 7218;
  Best Local Similarity   6.1%;  Pred. No. 0.18;
  Matches   25;  Conservative  213;  Mismatches  172;  Indels    0;  Gaps    0;

Qy         14 CAGGTATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTTAATGT 73
Db       1046 CAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1105

Qy         74 TTTTTCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCC 133
Db       1106 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1165

Qy        134 ATTCCAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAG 193
Db       1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225

Qy        194 ATTGAGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAG 253
Db       1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285

Qy        254 TTCTGTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACT 313
Db       1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345

Qy        314 AGGGTTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTG 373
Db       1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405

Qy        374 GATTAGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTT 423
Db       1406 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCT 1455


RESULT 11
AX344573
LOCUS       AX344573                 349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 24 from Patent WO0200932.
ACCESSION   AX344573
VERSION     AX344573.1  GI:18492459
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of known genetic parameters within the mhc
  JOURNAL   Patent: WO 0200932-A 24 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo
                     sapiens)~Original length of seq 1: 3.673778 <223>~split as
                     follows.~seq 01 0.000.001 TO 0.349.980~seq 02 0.300.001
                     649.980~seq 03 600.001 949.980~seq 04 900.001
                     1.249.980~seq 05 1.200.001 1.549.980~seq 06 1.500.001
                     1.849.980~seq 07 1.800.001 2.149.980~seq 08 2.100.001
                     2.449.980~seq 09 2.400.001 2.749.980~seq 10 2.700.001
                     3.049.980~seq 11 3.000.001 3.349.980~seq 12 3.300.001
                     3.649.980~seq 13 3.600.001 3.673.778 <223>~Original length
                     of seq 2: 3.673778 <223>~split as follows.~seq 14
                     0.000.001 TO 0.349.980~seq 15 0.300.001 649.980~seq 16
                     600.001 949.980~seq 17 900.001 1.249.980~seq 18 1.200.001
                     1.549.980~seq 19 1.500.001 1.849.980~seq 20 1.800.001
                     2.149.980~seq 21 2.100.001 2.449.980~seq 22 2.400.001
                     2.749.980~seq 23 2.700.001 3.049.980~seq 24 3.000.001
                     3.349.980~seq 25 3.300.001 3.649.980~seq 26 3.600.001
                     3.673.778"
ORIGIN

  Query Match             10.0%;  Score 49.6;  DB 6;  Length 349980;
  Best Local Similarity   49.6%;  Pred. No. 0.13;
  Matches  127;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTTAATGTTTTT 77
Db      54523 TGTGGTTTTAATGTTATTAGATAGTTTTTTTGAGATTTTTGGTTAGTTTTTTTTTTTTTT 54582

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db      54583 TTTTTTAGTGAATTTTTTATATTTTTTTTTTTTTAAGTTTTTTAAGAAATTTTTATTTTT 54642

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db      54643 GTTTTTTTAGTAGAAAGTTTTTTTTGTTGTTTTAATAAAAAAAGGAGATTAAGAGGTACG 54702
```

```
Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db      54703 AATTTTTTTAGTTTATTAGATGTAAGATATTTTGTTTGTTTTTTATTTTTTAAATTTTTT 54762

Qy        258 GTATAACCAGATTTGT 273
Db      54763 TTAATATTTAATTTAT 54778
```

RESULT 12
AX345007
LOCUS AX345007 17869 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 78 from Patent WO0200928.
ACCESSION AX345007
VERSION AX345007.1 GI:18492893
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 78 03-JAN-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .17869
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Query Match 9.9%; Score 49; DB 6; Length 17869;
Best Local Similarity 46.0%; Pred. No. 0.23;
Matches 210; Conservative 0; Mismatches 240; Indels 7; Gaps 1;

```
Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db      17051 TTTGTTTTTGTTTTTTTGAAAATTTATTGATAGTTTTTTATTTTTTTTTGTTGTTTTTTT 17110

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db      17111 TTGTTTGTTTATTTTTCGTTTTTTTTGGTTGTCGGTTGATTTTAGTTTGGTTGTGTATTT 17170

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db      17171 GTTTTTGGTTGTTTTTTTCGTGTTTTTGATTTTGTTTGTGTTTTATTTAGGTTTAGTTTT 17230

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db      17231 TGT-------TTTTTTTTGGTTATTTTTGTGTGTGTTTTTGTTTTTTTTGTTCGTATTTTT 17283

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db      17284 GGTTGTTTTTGTTAGTTTGTTTTTATTTTTATTTGTTTATTTGTTTGAAGAGATAGGTAG 17343

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db      17344 TTTTTTTGTTTTTTTTATATTTATTTGTGTTTTTTTGTTTTTGTGATTTTTTTTTTGGTTT 17403

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATG 437
Db      17404 TTTTGTTGTTTTTTTGTTTTTTTGTATTTTTGTTTGTTTTTTTGTTTATTGTTATTTATT 17463

Qy        438 AAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATG 474
Db      17464 TTTTTTTAGAGTTTTGTTTTTATTTTTTTTTTTTTTG 17500
```

RESULT 13
AX348307
LOCUS AX348307 17869 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 2 from Patent WO0202806.
ACCESSION AX348307
VERSION AX348307.1 GI:18614343
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Method and nucleic acids for pharmacogenomic methylation analysis
JOURNAL Patent: WO 0202806-A 2 10-JAN-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1. .17869
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

Query Match 9.9%; Score 49; DB 6; Length 17869;
Best Local Similarity 46.0%; Pred. No. 0.23;
Matches 210; Conservative 0; Mismatches 240; Indels 7; Gaps 1;

```
Qy         18 TATGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTT 77
Db      17051 TTTGTTTTTGTTTTTTTGAAAATTTATTGATAGTTTTTTATTTTTTTTTGTTGTTTTTTT 17110

Qy         78 TCTGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTC 137
Db      17111 TTGTTTGTTTATTTTTCGTTTTTTTTGGTTGTCGGTTGATTTTAGTTTGGTTGTGTATTT 17170

Qy        138 CAGTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTG 197
Db      17171 GTTTTTGGTTGTTTTTTTCGTGTTTTTGATTTTGTTTGTGTTTTATTTAGGTTTAGTTTT 17230

Qy        198 AGTCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCT 257
Db      17231 TGT-------TTTTTTTTGGTTATTTTTGTGTGTGTTTTTGTTTTTTTTGTTCGTATTTTT 17283

Qy        258 GTATAACCAGATTTGTCTGTGTGCGATTGTCATTACCTAGCCGTGTATCGAGAACTAGGG 317
Db      17284 GGTTGTTTTTGTTAGTTTGTTTTTATTTTTATTTGTTTATTTGTTTGAAGAGATAGGTAG 17343

Qy        318 TTTTCGAGTCAATTTTGCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGGATT 377
Db      17344 TTTTTTTGTTTTTTTTATATTTATTTGTGTTTTTTTGTTTTTGTGATTTTTTTTTTGGTTT 17403

Qy        378 AGGGTTTTAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATG 437
Db      17404 TTTTGTTGTTTTTTTGTTTTTTTGTATTTTTGTTTGTTTTTTTGTTTATTGTTATTTATT 17463

Qy        438 AAAATCCCGAATTGACTGTTCAATTTCTTGTTAAATG 474
Db      17464 TTTTTTTAGAGTTTTGTTTTTATTTTTTTTTTTTTG 17500
```

RESULT 14
AL953898/c
LOCUS AL953898 176162 bp DNA linear VRT 27-SEP-2003
DEFINITION Zebrafish DNA sequence from clone CH211-232K7 in linkage group 5,
complete sequence.
ACCESSION AL953898
VERSION AL953898.20 GI:35763627
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 176162)
AUTHORS Barker,D.
TITLE Direct Submission

```
  JOURNAL   Submitted (27-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 2003 this sequence version replaced gi:35209226.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-232K7
            is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .176162
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-232K7"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             9.9%;  Score 49;  DB 5;  Length 176162;
  Best Local Similarity   48.7%;  Pred. No. 0.19;
  Matches  133;  Conservative    0;  Mismatches  140;  Indels    0;  Gaps    0;

Qy         20 TGATTTGTTTGTAATTAGATCAGGGGTTTAGGTCTTTCCATTACTTTTTAATGTTTTTTC 79
              | |||  ||| | |||  ||       ||     | |||  |||      ||| | | ||||
Db      50213 TCATTGCTTTTTTATTTTATATTACTTTGCTTTTTTTTTATTTAACCTTATTTTATTTTG 50154

Qy         80 TGTTACTGTCTCCGCGATCTGATTTTACGACAATAGAGTTTCGGGTTTTGTCCCATTCCA 139
              |  | | | | |          |   | ||       ||    |||  |||| | |   |||  |
Db      50153 TTATTATTTTTTGTTATATTTTATCTTGTTTCATTTCATTTTAGGTTGTTTATTATTTTA 50094

Qy        140 GTTTGAAAATAAACGTCCGTCTTTTAAGTTTGCTGGATCGATAAACCTGTGAAGATTGAG 199
              ||| ||  | || || | | |||   ||||  ||  |   | |   | |    |  |||
Db      50093 TTTTTAATTTTAATGTACTTTTTCTTGTTTCATGTTTTATTTATTTTATCTTGCCTGAT 50034

Qy        200 TCTAGTCGATTTATTGGATGATCCATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGT 259
              | || |  |||| |    || |   || || | | || ||||| | | | | | |||
Db      50033 TTTACTTTATTTTTATTTTGTTATTTTATTTAACTTTATTTTCATACATTGTATTTCATA 49974

Qy        260 ATAACCAGATTTGTCTGTGTGCGATTGTCATTA 292
                | |    ||||  ||   |  |||   ||||
Db      49973 CTTATTTTATTTTACTTAATTTTATTCCTATTA 49941
```

```
RESULT 15
AE014820
LOCUS       AE014820              250029 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 14 section 5 of 13 of the
            complete sequence.
ACCESSION   AE014820 AE014187
VERSION     AE014820.1  GI:23497305
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 250029)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 250029)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .250029
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="14"
     repeat_region   83. .122
                     /rpt_type=tandem
     repeat_region   307. .346
                     /rpt_type=tandem
     repeat_region   346. .442
                     /rpt_type=tandem
     repeat_region   575. .646
                     /rpt_type=tandem
     gene            <833. .>3733
                     /locus_tag="PF14_0238"
     mRNA            <833. .>3733
                     /locus_tag="PF14_0238"
     CDS             833. .3733
                     /locus_tag="PF14_0238"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN36851.1"
                     /db_xref="GI:23497306"
                     /translation="MNNDIKNELINNLPFKFSDISLNSFEKKKEKDKDYIPDEDGGHT
                     WGDFSRDKIGNVKKKEKQKKKKDNEDYSDDNNDNNNNYSSDHNNNNNNSSSDHNNNKR
                     LVLLKEKLKYNHYYYAGKLAEKEWEENYDNLKKKSQLFKDVLEKDKGKNFSTFNITKN
                     EKICNIKEKAQKKKQNKNQKNLKKKNFKKEHNDISFNDTYTKYSSSLNDFNDISDSLN
                     LKNDIINNEEEYNLTNSLFQSFPMEQNLPLFKYKKENKQDYDNNVKCREYDEKKVPKG
                     IIHEIIEGEEEEEGENKKKKKKKKKKNTLDIVHNVINKKISYDDNVNIPYFEDDDNKT
                     NILGENMSNIFEEDKELDLFSSFKDIFNNHKKKIKLIDYNKILNKSLYDNVENVDSLI
                     AQEKRKKGKKEERKPQKKKKKKRNNDQNSAHNEENMKQIKNNDNFNNMQSLNEYNEKV
                     RGNFFEKKKTENDVKEIIKNIKMRLGFYSNEESDSHDIKCLENDDDNDDDDNNDDNDD
                     NNDDNDDNNDDDDNNNDGDDDNNNNNDDDNNKNDIYNNLIINEEEEKQNLENLKDNIN
                     IPYTSKELYTYKEYSKKLDCNDLNKLSDEKETTKNISNDLELIKKEKKKKNDDTFFFI
                     SHNEKDKNYLFPYDHKGDDNYNDEDKNNLLPCDHICDDNYNDEDKNNLLPCDHICDDN
                     YNDEDLFSLSMSSITFDKRKGIKNMECKKSIECVDDKNKLSINYDENNNKNILSYNIN
                     VQDENDKLYKNEQKIKQDIYNDNENDKLYKNEQKIKQDIYNDNENDKLYKNDQKIKQD
                     IYNDNENDKLYKNDQKIKQDIYNDNENDMLINSMSPENTKYIKNIKENDKLLKNVMKK
                     EEIDKKEKSNICICKYIDKYDMKNLIDEIKKKKKKNDVENSNNSYSEDDVCFYELYKF
                     ICSNNIEGNNENIFDDTNINVENVSSNIDNYMQSIFNNQIKYLDNLISQIFS"
```

```
repeat_region   1704. .1740
                /rpt_type=tandem
repeat_region   2309. .2446
                /rpt_type=tandem
repeat_region   2310. .2443
                /rpt_type=tandem
repeat_region   2313. .2449
                /rpt_type=tandem
repeat_region   2718. .2893
                /rpt_type=tandem
repeat_region   3065. .3311
                /rpt_type=tandem
repeat_region   3065. .3273
                /rpt_type=tandem
repeat_region   3776. .3805
                /rpt_type=tandem
repeat_region   4039. .4094
                /rpt_type=tandem
repeat_region   4260. .4295
                /rpt_type=tandem
repeat_region   4402. .4439
                /rpt_type=tandem
gene            complement(<4519. .>6088)
                /locus_tag="PF14_0239"
mRNA            complement(join(<4519. .4588,4895. .5573,5806. .>6088))
                /locus_tag="PF14_0239"
CDS             complement(join(4519. .4588,4895. .5573,5806. .6088))
                /locus_tag="PF14_0239"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAN36852.1"
                /db_xref="GI:23497307"
                /translation="MEEQNEEIEKTVEESEEINEEVYEELFLFASSDNDLVKKESLRI
                ILSLLEEKEFIDHIIKNNKKYLKTLISSLNSRCKLLTLECLLNLSAQLPKELTERNLI
                EILFDMMKEEEKIEDNCIDLYIMIISNLTRCKEGVYKVLDINNDSNINIKEDNFKVSF
                FLNKLLYFFFLPIKPSINKNLSDKYIYVSHVLINISSIKESIVFFKNVAFLNKISDQI
                LNVERFRAILPFIINLCLNEVIHPYIFHDDCYLFPYVLSYLYTNDYNITKSGSYNNSN
                NNEEINTQNIHHIIMNKSSILVSCSVIKSRILIILFYLCNRDYSRNDIENVTNKLIEG
                SNSREAIAT"
repeat_region   4677. .4714
                /rpt_type=tandem
repeat_region   5600. .5641
                /rpt_type=tandem
repeat_region   5745. .5791
                /rpt_type=tandem
repeat_region   6108. .6152
                /rpt_type=tandem
repeat_region   6168. .6222
                /rpt_type=tandem
repeat_region   6220. .6268
                /rpt_type=tandem
repeat_region   6420. .6481
                /rpt_type=tandem
repeat_region   6563. .6621
                /rpt_type=tandem
repeat_region   6718. .6780
                /rpt_type=tandem
repeat_region   6719. .6790
                /rpt_type=tandem
repeat_region   7543. .7568
                /rpt_type=tandem
repeat_region   7761. .7816
                /rpt_type=tandem
repeat_region   7824. .7849
                /rpt_type=tandem
repeat_region   7871. .7930
                /rpt_type=tandem
gene            complement(<7953. .>8438)
                /locus_tag="PF14_0240"
mRNA            complement(<7953. .>8438)
                /locus_tag="PF14_0240"
CDS             complement(7953. .8438)
                /locus_tag="PF14_0240"
                /codon_start=1
                /product="ribosomal protein L21e, putative"
                /protein_id="AAN36853.1"
                /db_xref="GI:23497308"
                /translation="MGGKSRGKRSGTRYKFSKKFRKHGECTANKYLEKLNYGDYVDIV
                CDSTQQKGMPFNYYHGKTGKIFHITKRGVGVLVNKRVKHRIEQKKVCVRIEHVRKSRC
                NEDFLLRKIKNAELIKEAKLKNEHINIKRKTEGPKPAAMIKVPPSKIITIEPLPFYEE
                Y"
repeat_region   8467. .8550
                /rpt_type=tandem
repeat_region   8473. .8539
                /rpt_type=tandem
repeat_region   8477. .8537
                /rpt_type=tandem
repeat_region   8723. .8844
                /rpt_type=tandem
repeat_region   8723. .8840
                /rpt_type=tandem
repeat_region   8782. .8844
                /rpt_type=tandem
repeat_region   9019. .9059
                /rpt_type=tandem
repeat_region   9019. .9058
                /rpt_type=tandem
repeat_region   9133. .9189
                /rpt_type=tandem
repeat_region   9221. .9254
                /rpt_type=tandem
repeat_region   9604. .9637
                /rpt_type=tandem
repeat_region   9668. .9723
                /rpt_type=tandem
repeat_region   10036. .10073
                /rpt_type=tandem
repeat_region   10445. .10472
                /rpt_type=tandem
repeat_region   10511. .10551
                /rpt_type=tandem
gene            complement(<10565. .>11080)
                /locus_tag="PF14_0241"
mRNA            complement(<10565. .>11080)
                /locus_tag="PF14_0241"
CDS             complement(10565. .11080)
                /locus_tag="PF14_0241"
                /codon_start=1
                /product="basic transcription factor 3b, putative"
                /protein_id="AAN36854.1"
                /db_xref="GI:23497309"
                /translation="MEKISPEILAARAKLKEKMGGNLRQIGGKGSARRKIKKVHKNSI
                SNEKKINIILKKIGASYFGDVDEICVYRTGDTFLEFKKPKLCASLQSNTYIVTGKFNE
                HKIDINKLFEGLKGNKNLDMNLLEKIKNDPNIKNILNKESGDTPKREEDEQEANDVPD
                LVENFEEVSKE"
repeat_region   11429. .11458
                /rpt_type=tandem
repeat_region   11514. .11584
                /rpt_type=tandem
repeat_region   11514. .11579
```

```
Query Match             9.8%;  Score 48.6;  DB 3;  Length 250029;
Best Local Similarity   21.8%;  Pred. No. 0.22;
Matches   57;  Conservative   95;  Mismatches  109;  Indels    0;  Gaps    0;

Qy        223 CATTCTTCATCGTTTTTTTCTTGCTTCGAAGTTCTGTATAACCAGATTTGTCTGTGTGCG 282
              |||  || ||  | |  || |   ||: ||  |   |:|::  : ::|: | : : ::
Db     129266 CATCATTTATATTATCATTATAATTTYCAATATTATTWTWWTWWTWWTWWTWWYSKY 129325

Qy        283 ATTGTCATTACCTAGCCGTGTATCGAGAACTAGGGTTTTCGAGTCAATTTTGCCCCTTTT 342
              ::::: ::|:  : : : |  |::   :: | ::::::::|:  |:::||::  ::   ::
Db     129326 MWKRWTWWTWWWWKKWMYTWAAWYYTTWWTTYSKKKWWYCSKTTYMWTTWWWWYMARRWY 129385

Qy        343 GGTTATATCTGGTTCGATAACGATTCATCTGGATTAGGGTTTTAAGTGGTGACGTTTAGT 402
                : :: : |  :|:  :  : ::   : :: :|:|    :: |: :  : :: :||: |
Db     129386 YTYMWYYWWTTYYTYYTYYTYMWYMWKKTKKWWTWAAWYCYYMAWWWYMWYMMCYTTWWT 129445
```

```
Qy        403 ATTCCAATTTCTTCAAAATTTAGTTATGGATAATGAAAATCCCGAATTGACTGTTCAATT 462
              |:   ::|::|| :   ::|: ::|  : ::::  :||:   |:::|  |     | :|
Db     129446 TTKRWTWWTYYTTWWTTTKKTWAWWAAWKKWWWWYCMAAWTTKGRWKTCTCMMMAAAGKT 129505

Qy        463 TCTTGTTAAATGCGCAGATCC 483
              |:  : ||: | : :|  | |
Db     129506 TYCMRCTAWTTYMWYATCTTC 129526
```

Search completed: June 20, 2004, 03:45:49
Job time : 3467.42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:40:43 ; Search time 59.125 Seconds
(without alignments)
3923.374 Million cell updates/sec

Title: US-09-000-062-6
Perfect score: 418
Sequence: 1 tgaggtacgattcttcgatc.....taatttgttgaacagatccc 418

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	418	100.0	418	4	US-09-000-062-6	Sequence 6, Appli
	2	43.8	10.5	5852	1	US-07-867-106-2	Sequence 2, Appli
	3	42.4	10.1	8537	4	US-10-204-708-41	Sequence 41, Appl
	4	40.2	9.6	5455	4	US-10-204-708-34	Sequence 34, Appl
	5	40.2	9.6	11049	4	US-10-204-708-23	Sequence 23, Appl
	6	39.6	9.5	11015	4	US-10-204-708-55	Sequence 55, Appl
	7	39.4	9.4	5340	4	US-09-627-122-21	Sequence 21, Appl
	8	39.2	9.4	11050	4	US-10-204-708-86	Sequence 86, Appl
	9	39	9.3	5501	4	US-10-204-708-37	Sequence 37, Appl
	10	38.4	9.2	5562	4	US-10-204-708-63	Sequence 63, Appl
	11	38.2	9.1	5501	4	US-10-204-708-38	Sequence 38, Appl
	12	38.2	9.1	6801	4	US-10-204-708-61	Sequence 61, Appl
	13	38.2	9.1	6866	4	US-10-204-708-20	Sequence 20, Appl
c	14	38	9.1	786431	4	US-09-751-389-3	Sequence 3, Appli
	15	37.8	9.0	5455	4	US-10-204-708-33	Sequence 33, Appl
	16	37.8	9.0	6801	4	US-10-204-708-62	Sequence 62, Appl
c	17	37.8	9.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
	18	37.6	9.0	11050	4	US-10-204-708-85	Sequence 85, Appl
	19	37.6	9.0	11811	3	US-09-078-294-7	Sequence 7, Appli
	20	37.4	8.9	5666	4	US-10-204-708-29	Sequence 29, Appl
	21	37.4	8.9	6306	4	US-10-204-708-50	Sequence 50, Appl
	22	37.4	8.9	23439	4	US-08-956-171E-38	Sequence 38, Appl
	23	37.2	8.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
	24	37	8.9	6317	4	US-10-204-708-11	Sequence 11, Appl
c	25	37	8.9	10640	4	US-09-417-485D-5	Sequence 5, Appli
c	26	36.6	8.8	984	4	US-09-134-001C-2100	Sequence 2100, Ap
	27	36.6	8.8	19513	4	US-10-204-708-40	Sequence 40, Appl
c	28	36.6	8.8	640681	4	US-09-790-988-1	Sequence 1, Appli
	29	36.4	8.7	6040	4	US-10-204-708-70	Sequence 70, Appl
	30	36.4	8.7	6182	4	US-10-204-708-88	Sequence 88, Appl
	31	36.4	8.7	8607	4	US-10-204-708-71	Sequence 71, Appl
	32	36.4	8.7	202001	4	US-09-734-674-3	Sequence 3, Appli
	33	36.2	8.7	6113	4	US-10-204-708-14	Sequence 14, Appl
	34	36.2	8.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	35	36.2	8.7	8093	4	US-10-204-708-32	Sequence 32, Appl
	36	36.2	8.7	11015	4	US-10-204-708-56	Sequence 56, Appl
	37	36	8.6	6070	4	US-10-204-708-10	Sequence 10, Appl
	38	36	8.6	19233	4	US-10-204-708-45	Sequence 45, Appl
c	39	35.8	8.6	2103	4	US-09-543-681A-797	Sequence 797, App
	40	35.8	8.6	5666	4	US-10-204-708-30	Sequence 30, Appl
	41	35.8	8.6	8607	4	US-10-204-708-72	Sequence 72, Appl
	42	35.4	8.5	522	4	US-09-134-001C-1017	Sequence 1017, Ap
	43	35.4	8.5	8961	4	US-10-204-708-80	Sequence 80, Appl
c	44	35.2	8.4	6243	2	US-09-056-075-1	Sequence 1, Appli
	45	35.2	8.4	8537	4	US-10-204-708-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1
US-09-000-062-6
; Sequence 6, Application US/09000062
; Patent No. 6338961
; GENERAL INFORMATION:
; APPLICANT: DEROSE, Richard
; APPLICANT: CHAUBET, Nicole
; APPLICANT: GIGOT, Claude (deceased)
; TITLE OF INVENTION: ISOLATED DNA SEQUENCE CAPABLE OF SERVING AS REGULATORY
; TITLE OF INVENTION: ELEMENT IN A CHIMERIC GENE WHICH CAN BE USED FOR THE
; TITLE OF INVENTION: TRANSFORMATION OF PLANTS
; FILE REFERENCE: 022650-453
; CURRENT APPLICATION NUMBER: US/09/000,062
; CURRENT FILING DATE: 1998-05-29
; EARLIER APPLICATION NUMBER: PCT/FR96/01109
; EARLIER FILING DATE: 1996-07-17
; EARLIER APPLICATION NUMBER: FR 95/08980
; EARLIER FILING DATE: 1995-07-19
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Zea mays
US-09-000-062-6

Query Match 100.0%; Score 418; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 1.8e-102;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGAGGTACGATTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGA 60

Qy         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AACTACTGGAATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGG 120

Qy        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAA 180

Qy        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGA 240

Qy        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGAT 300
```

```
Qy        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATA 360

Qy        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CATTCCTGCAATTGAATACGTATGGATCTAAATCTTGTTAATTTGTTGAACAGATCCC 418


RESULT 2
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2

  Query Match             10.5%;  Score 43.8;  DB 1;  Length 5852;
  Best Local Similarity   50.7%;  Pred. No. 0.014;
  Matches  105;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy        201 ACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAAT 260
              |||  ||    |   |  |  | ||| | | ||||  |    ||||||  ||||   |  |
Db       2126 ACCTTATTTATAAATATTGGATAATATATCAAATATTTATCAGTTTTGGCATGACAATTT 2185

Qy        261 AGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGC 320
              |||  |  |||||    ||||  |  ||    ||||  ||||  |  |  |  |    ||
Db       2186 TAATTATATTTATTTTTTGATTAATTTTTTTTTTTTTTTTTTTTTAAAATTTCTTTTTTT 2245

Qy        321 AATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACG 380
              ||    ||||  |  ||  | |||  | || | | || |||  |   |||  ||
Db       2246 TTTTTTTATTTTTAATTTTTAATTTTTATTTTTCCCACACTTTCATTTTATTTTATTTT 2305

Qy        381 TATGGATCTAAATCTTGTTAATTTGTT 407
              |   | |||||    || |||| ||
Db       2306 ATTTATTGTAAATTCATTTTATTTATT 2332


RESULT 3
US-10-204-708-41
; Sequence 41, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 41
;   LENGTH: 8537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-41

  Query Match             10.1%;  Score 42.4;  DB 4;  Length 8537;
  Best Local Similarity   50.5%;  Pred. No. 0.036;
  Matches  103;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        143 AATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCAC 202
              | |  | ||   | ||  ||||||   |  |   ||||  || |  |   |||  | |
Db       7785 ATTAGTGTTTATTGTAGTTATTTTATTGTGAAATAGAATTTTAGAATTTATTTTTTTTAT 7844

Qy        203 CAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAG 262
              || |   |||||  |     |   | |  || |   |  || ||   | |   |||||
Db       7845 TTAATTGAAATTTTGTGTTTGTTGATTAATTTTTTTTTTTTTTTGTCGTTTTTTGTAATAG 7904

Qy        263 ATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAA 322
              |   | || ||  |   ||| | ||||  |||||  ||  | ||   ||||  |||
Db       7905 TTGTATTTTTTTTAGTTTTTGGTTATTATTGTTTTATTTGTTTTTTGTATGAGATGAATT 7964

Qy        323 TTAGGGATTTTCTTTGGTTTTGTG 346
              ||    |||| |||  |||| ||
Db       7965 TTTTTTTTTTTGTTTTTTTTTTTG 7988


RESULT 4
US-10-204-708-34
; Sequence 34, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
```

```
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 34
;   LENGTH: 5455
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-34

  Query Match            9.6%;  Score 40.2;  DB 4;  Length 5455;
  Best Local Similarity  47.8%;  Pred. No. 0.12;
  Matches  117;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy        104 TTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAGAT 163
Db       2998 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTCGTTTTTATTG 3057

Qy        164 TTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGA 223
Db       3058 TTTAGGTTGGAGTGTAATGGTATTATTTTAGTTTATTGTAATTTTTATTTTTCGGGTTTA 3117

Qy        224 ATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTG 283
Db       3118 ATTGATTTTTTTGTTTTAGTTTTTTAAGTAGTTGGGATTATAGGTGTTTATTATTTCGTT 3177

Qy        284 TTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTT 343
Db       3178 TGGTAAATTTTTGTATTTTTAGTAGAGATAGGGTTTTATTATGTTGGTTAGTTTGGTTTT 3237

Qy        344 GTGTT 348
Db       3238 GATTT 3242

RESULT 5
US-10-204-708-23
; Sequence 23, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 23
;   LENGTH: 11049
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-23

  Query Match            9.6%;  Score 40.2;  DB 4;  Length 11049;
  Best Local Similarity  43.8%;  Pred. No. 0.15;
  Matches  174;  Conservative    0;  Mismatches  223;  Indels    0;  Gaps    0;

Qy         11 TTCTTCGATCCTCTTTGATTTTCCTGGAAATATTTTTTCGGTGATCGTGAAACTACTGGA 70
Db       7090 TTTTTTATTTTTAGGTTTTTTTTGTTTTTTTGTTTGTTTGTTTTTGTTTTTTTAAATTTA 7149

Qy         71 ATCGCTCGATAGGTGGTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTT 130
Db       7150 TTTAGTTTGGATTTTGAAGTTAGTTGTTTTAGTTATTTTTTTTAGTATTTTTAATTTCGT 7209

Qy        131 GTTTCTTCGCCGAATGATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCT 190
Db       7210 GATTTTTTTGCGTTATTATTTTTGTTTTGAAAATTTTTAATAGTGATATTTTTTTTAGGA 7269

Qy        191 AGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCAC 250
Db       7270 TTATTGAAGGAGGAGATTGTTTAGTTTTGGAGATTGTTAGTAGTTTATGGTTTTGTAGGG 7329

Qy        251 ATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATT 310
Db       7330 TTGGGGAGATTTTTAGGTAGGTTTTTATAGTAATTATTAGTGTTTGTTTTTTGTTAATT 7389

Qy        311 ATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCA 370
Db       7390 TTTTTTTTTTTTGTATAGGATATTTATTTTAGATTTTTAGTTTTTAGTTTTAGTGTTAAT 7449

Qy        371 ATTGAATACGTATGGATCTAAATCTTGTTAATTTGTT 407
Db       7450 TTTGTTTTTTGATTTAGGAATTGTTTAATTTGTTGTT 7486

RESULT 6
US-10-204-708-55
; Sequence 55, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 55
;   LENGTH: 11015
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-55

  Query Match            9.5%;  Score 39.6;  DB 4;  Length 11015;
  Best Local Similarity  54.9%;  Pred. No. 0.21;
```

```
  Matches   78;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy        209 GTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTT 268
              |||| ||||| |     | |    || |||||  || | |  | | |||| |||  |||
Db       3003 GTTTTTTTGTTTTTTTATTATTTTGTTATCGATTTTTATAAAGTTTTTAAAGAGAAAGTT 3062

Qy        269 CTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGG 328
              |  |   |  |||||    | |  ||| |||  ||||||          | | ||  |
Db       3063 TGTTATATGTTGTTGTTTTAGAAGTGTTTTATTAATTTGATTGGTTAGGAGGTAATATAG 3122

Qy        329 ATTTTCTTTGGTTTTGTGTTGA 350
              |||| ||||  | ||||||
Db       3123 GTTTTTTTTGAATATGTGTTTA 3144


RESULT 7
US-09-627-122-21
; Sequence 21, Application US/09627122
; Patent No. 6472521
; GENERAL INFORMATION:
;  APPLICANT:  Uhlmann, Eugen
;  APPLICANT:  Greiner, Beate
;  APPLICANT:  Unger, Eberhard
;  APPLICANT:  Gothe, Gislinde
;  APPLICANT:  Schwerdel, Marc
;  TITLE OF INVENTION:  OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN eg5
;  TITLE OF INVENTION:  EXPRESSION
;  FILE REFERENCE:  02481.1678
;  CURRENT APPLICATION NUMBER: US/09/627,122
;  CURRENT FILING DATE:  2000-07-27
;  NUMBER OF SEQ ID NOS:  23
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 5340
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-627-122-21

  Query Match             9.4%;  Score 39.4;  DB 4;  Length 5340;
  Best Local Similarity   51.3%;  Pred. No. 0.2;
  Matches  143;  Conservative    0;  Mismatches  131;  Indels    5;  Gaps    2;

Qy         86 GTACGAAATTAGGCGAGATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAAT 145
              | |  | |||||  |     ||  |||    || |||     |  | || ||  ||      |||
Db       2960 GCATTATATTAGACATATTTGTTTTAATTTGTTGAAAAGATTTTTTTTGATTTATAAAAT 3019

Qy        146 GATCTTCCGTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAG 205
              |||||    | ||  ||| ||    |||     | |  |      |  ||  ||||  |   |
Db       3020 TTTCTTCTAAAGAATGATATTTATTTAATAAGAATTGTGTAATATATTTTTCTT---CTA 3076

Qy        206 ATAGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATT 265
              || ||| ||  |  | | |    ||||||  || |     ||||  | ||  | | |||||  |
Db       3077 TTATTTTTTTAATTAATATTTGATGAAATGCTTGTATATTTTTATATTTTTGAATAGTAT 3136

Qy        266 GTTCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTA 325
              || | |   |   | | ||  |||    | | |||  |||  |||||  |  ||    | |
Db       3137 CTTTTAAAAAGAAAAATATT--TTATTTTGTAGATCATCTGTATTATCCATTTTATTTAA 3194

Qy        326 GGGATTTTCTTTGGTTTTGTGTTGATCTTACGATACATT 364
              |||||  |  ||||   ||  ||  |  |  ||
Db       3195 TAAATTTTTTATTTTTTTACTTTTTTCAAATAAAATTTT 3233


RESULT 8
US-10-204-708-86
; Sequence 86, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 86
;   LENGTH: 11050
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-86

  Query Match             9.4%;  Score 39.2;  DB 4;  Length 11050;
  Best Local Similarity   46.9%;  Pred. No. 0.27;
  Matches  122;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy        154 GTATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTC 213
              ||| | |      ||  | || |||    |  ||| | ||     |  |          ||||
Db        268 GTAAATATTCGATAAATGAGTGATTTTTTTTATTGTTATTAGTATTTTGATATGGGTTTG 327

Qy        214 TTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTAT 273
              | || |  |  | |   | ||| |    || |||      ||| || ||    |  | || |
Db        328 TATGACGTGGTTTTGGTAGATTTTTTTTATTTTATTTTTGTTTATATTTTATCGTTTTTT 387

Qy        274 TCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTT 333
              |  |  ||  | ||   ||||||  |  ||   ||  |    |||   |  ||  |||
Db        388 TGAGTTTTTAATTATATTGGTTTTTAGGTTTGATATATTAATTTTGTTTTAGGGTTTTTG 447

Qy        334 CTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAAT 393
              ||||||  | | ||     ||   |||  ||  |       |  ||  |   | | |   |
Db        448 AATTGGTTGTTTTTTTTTATTTGGAATATTTTTTTTTTGAGGTTTATATGGTGGTTTTTTT 507

Qy        394 CTTGTTAATTTGTTGAACAG 413
              | |||| || |||  | ||
Db        508 TTAGTTATTTAGTTTTAGAG 527


RESULT 9
US-10-204-708-37
; Sequence 37, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 37
;   LENGTH: 5501
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-37

  Query Match             9.3%;  Score 39;  DB 4;  Length 5501;
  Best Local Similarity   49.8%;  Pred. No. 0.26;
  Matches  126;  Conservative    0;  Mismatches  125;  Indels    2;  Gaps    1;

Qy        102 GATTAGTTTCTATTCTTGGCCATTATCTTGTTTCTTCGCCGAATGATCTTCCGTATAAAG 161
              | | |  |  |||| ||    ||| | | | || || |   |||| | |   || ||| |
Db       1246 GTTGATGTAATATTTTTAAAAATTTTGTGGGTTTTTTGTGTAATGTTATGGAGTTTAATG 1305

Qy        162 ATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTA 221
              | |||| | ||||      |     |  |   |  || |      || |||  || |  |
Db       1306 TTATTAGATAAGAGGGTTTTTTATAAATTTTTTGGGATAA--TTTTATTTTTATTTTTGA 1363

Qy        222 GAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGAT 281
              |   ||| ||  | ||  |  || | | |     |  || ||  |
Db       1364 TTTTTATTGAGATGGTTGAATGAGTTTAAGAGTTTTATGTTTTATTTTTCGTGTTATTA 1423

Qy        282 TGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTT 341
              | || | ||   |||| | || |||   | | ||| |   || |  || || ||| |||
Db       1424 TATTTAGTATATTTTTTTTATTTTTGTAAAAGAGATGGGGTTTCGTTATGTTGTTTAGTT 1483

Qy        342 TTGTGTTGATCTT 354
              | || ||||  ||
Db       1484 TGGTTTTGAATTT 1496


RESULT 10
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
;   LENGTH: 5562
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

  Query Match             9.2%;  Score 38.4;  DB 4;  Length 5562;
  Best Local Similarity   49.5%;  Pred. No. 0.37;
  Matches   99;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy        208 AGTTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGT 267
              ||||| || |  |    | |     ||| || | |||     || |   | |  ||| |
Db        367 AGTTTTTTAGGTTTTTTTTTAATTTATTAACGTTGGTTGGGTGATTTTCGATTTTATTTT 426

Qy        268 TCTTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGG 327
              | || |   |   || | |||||| |||||  |||| |   | | |   ||   |||

Db        427 TTTTTTATAGTTTTTTTCGATTAGTGTTTTTTTTTTGTATTTAAAGTCGTTTTTATTTTT 486

Qy        328 GATTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGAT 387
              ||  ||  ||  | ||  | ||  |||  |   |    ||  |  ||||  ||
Db        487 TTATTAGGTTTTATTATTTTTTTTTTTGTGATTGAAAGTTTTTTTTAGTTTTGTATAAAT 546

Qy        388 CTAAATCTTGTTAATTTGTT 407
              |   | || || ||| |||
Db        547 ATTTTTGTTTTTTATTGGTT 566


RESULT 11
US-10-204-708-38
; Sequence 38, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 38
;   LENGTH: 5501
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-38

  Query Match             9.1%;  Score 38.2;  DB 4;  Length 5501;
  Best Local Similarity   49.7%;  Pred. No. 0.42;
  Matches   97;  Conservative    0;  Mismatches   98;  Indels    0;  Gaps    0;

Qy        210 TTTCTTTGTCTAGAATCTCTGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTC 269
              | | ||||| ||| ||    |   |  | || |||||    |  ||   |  || ||
Db       3630 TGTGTTTGTTTAGTATTGTTTTTTTATTGTATGGTTTTTTTTTTGGTTGTTTTTTTTTTA 3689

Qy        270 TTATTCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGA 329
              |||||  |  ||  |   | | |  |||  |  || |  ||||   |||||
Db       3690 TTATTTAGGTTTTAGTTTGGTAGTTATATTTTTGTTACGTTTTTGGATTAGGTTTAGGTT 3749

Qy        330 TTTTCTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCT 389
              |||  ||  ||  || | || ||  ||  |   |   || || |   ||   ||| | |
Db       3750 TTTATTTGTGTGTTTTTTTTATAGTATTAAGAAGGATTGGAAATATTTATACGTGGTTGT 3809

Qy        390 AAATCTTGTTAATTT 404
              | ||  |||   |||
Db       3810 ATATTATGTATTTTT 3824


RESULT 12
US-10-204-708-61
; Sequence 61, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
```

```
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 61
;   LENGTH: 6801
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-61

  Query Match           9.1%;  Score 38.2;  DB 4;  Length 6801;
  Best Local Similarity   50.8%;  Pred. No. 0.45;
  Matches   91;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        170 TTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCTTTGTCTAGAATCTCT 229
              | || ||| ||| || | | | |       | || | | | ||| | | ||| | |
Db        192 TGAGGGATAAATTGTTTTTCGTGGTAGTTGTATTAGTTTATATTTTTATTAATAATGTAT 251

Qy        230 GAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATT 289
              ||    | |  ||  |||| | || | |     | ||| || || ||       || ||| ||
Db        252 GAGGGTTTTAATTTTTTTTATATTTTTGTTAATATTTTTTTTTTTTTTTTTTTTTTGTTT 311

Qy        290 AGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTT 348
                |||  |||||  | |||||       | |        ||  |||| ||  || ||
Db        312 TTTTTTAAATTTTAGTTATTATTATTAAAGATTATAAAGGTTATTTTGATTAAGTTTT 370


RESULT 13
US-10-204-708-20
; Sequence 20, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 20
;   LENGTH: 6866
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-20

  Query Match           9.1%;  Score 38.2;  DB 4;  Length 6866;
  Best Local Similarity   49.4%;  Pred. No. 0.45;
  Matches  126;  Conservative    0;  Mismatches  128;  Indels    1;  Gaps    1;

Qy        155 TATAAAGATTTTAGGTTAGAGATGAATCGTATAGCTAGATTTCATCACCAGATAGTTTCT 214
              |||  |||||||   || || ||   | | |   | |       | | |  | |||
Db       1005 TATTTAGATTTTTTTTTCGATATTTTTTTTTTTTTTGTAGCGATTTTCGATTTTTTTTA 1064

Qy        215 TTGTCTAGAATCTCTGAAATTCTCGATAGTTTT-CACATGTGTAAATAGATTGTTCTTAT 273
              || | |||| | | |  | || |    | |||  |  | ||| |      ||||  || |
Db       1065 TTATTTAGAGTTTTTTTATTTTTAGTGTTTTTGTATTTATGTTAGGGATTTGTATTTTT 1124

Qy        274 TCGGCGATTGTTGATTAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTT 333
              | || |||| ||  |      |||| | || ||     |  | |  |     || |||||||||
Db       1125 TTGGGGATTTTTTTTATTTTTTTTAGTGTTGTTAGGGAATTGTATTTTTTTCGGGATTTT 1184

Qy        334 CTTTGGTTTTGTGTTGATCTTACGATACATTCCTGCAATTGAATACGTATGGATCTAAAT 393
              |||  | ||| |  | | ||   |||   |  | | | | ||  | ||  |  |
Db       1185 TTTTATTATTTTTATTGTGTTTGTATAGTGTTAGGGATTCGTATTTTTTGGGGATTTTTT 1244

Qy        394 CTTGTTAATTTGTTG 408
              |  ||  |||  ||
Db       1245 TTATTTTTTTTAGTG 1259


RESULT 14
US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
;  APPLICANT: GUEGLER, Karl et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001067
;  CURRENT APPLICATION NUMBER: US/09/751,389
;  CURRENT FILING DATE:  2001-01-02
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 786431
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(786431)
;   OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

  Query Match           9.1%;  Score 38;  DB 4;  Length 786431;
  Best Local Similarity   55.2%;  Pred. No. 1.8;
  Matches   74;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        229 TGAAATTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGAT 288
              || ||   ||  ||  ||||  || | || |   |||||  |||      ||| ||  |
Db     455675 TGGAAAATTCATTATATTTCCAATCTCTATAATATTTGTTACTATATCTTCATTTTTCTT 455616

Qy        289 TAGGGTTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTT 348
                |||||  | |||||||  || |   |||    | |    || |||||  |||| |
Db     455615 CCCAGTTTTTAGTTTCTTGTTTCTTTTATTTTGGTCAATTTTTATCTTTTTTTTTTTAAG 455556

Qy        349 GATCTTACGATACA 362
              |||  |  |||| |
Db     455555 AATCAAAGGATAAA 455542


RESULT 15
US-10-204-708-33
; Sequence 33, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
```

; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 33
; LENGTH: 5455
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-33

Query Match 9.0%; Score 37.8; DB 4; Length 5455;
Best Local Similarity 53.8%; Pred. No. 0.54;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy        234 TTCTCGATAGTTTTCACATGTGTAAATAGATTGTTCTTATTCGGCGATTGTTGATTAGGG 293
              || |   |  |||     | | |   ||  || || || ||  | | ||| | | |  |
Db       3679 TTTTTTTTTTTTTGGGTTTTTTTTTTATTTTTTTTTTTTTTGGGTTTGGTAAATTAGT 3738

Qy        294 TTTTGATTTTCTTGATTATGCGATTGCAATTAGGGATTTTCTTTGGTTTTGTGTTGATCT 353
              |||| ||||| ||   | |  ||||| ||   ||| |||| |||| | || | | ||  |
Db       3739 TTTTAATTTTTTTTATATTTTAGATTGGAAGGTGGGTTTTTTTTTGATATTTTTTAGAGTT 3798

Qy        354 TACGATACATTCCTGCAATTGAATA 378
                ||  ||  |  ||  |  ||
Db       3799 GTTATTAATTTTTTTTAAGTTTTTA 3823
```

Search completed: June 20, 2004, 05:03:40
Job time : 61.125 secs